EXHIBIT A

Biochemistry 1993, *32*, 6250–6258

Random Mutagenesis of the Substrate-Binding Site of a Serine Protease Can Generate Enzymes with Increased Activities and Altered Primary Specificities†

Lloyd D. Graham,* Kevin D. Haggett, Philip A. Jennings,‡ Darren S. Le Brocque, and Robert G. Whittaker

C.S.I.R.O. Division of Biomolecular Engineering, P.O. Box 184, North Ryde, New South Wales 2113, Australia

Paul A. Schober

Peptide Technology Ltd., P.O. Box 444, Dee Why, New South Wales 2099, Australia

Received October 30, 1992; Revised Manuscript Received March 25, 1993

ABSTRACT: In the past, several point mutations have been introduced individually into the substrate-binding site of α-lytic protease (EC 3.4.21.12) and shown to affect its specificity in a predictable manner [Bone, R., Silen, J. L., & Agard, D. A. (1989) *Nature 339*, 191–195]. One of the resulting mutant enzymes (Met190Ala in the numbering system of Fujinaga et al.) [Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1985) *J. Mol. Biol. 183*, 479–502] cleaves at large hydrophobic residues. We chose this enzyme as the parent for a library of mutant proteases. The library was constructed by effecting combinatorial random substitution of up to four other residues (Gly191, Arg192, Met213, and Val218) thought likely to influence the primary specificity of the protease. Active enzymes in the library were screened with a range of synthetic substrates (encompassing 19 different amino acids in the P_1 position) in order to evaluate their primary cleavage preferences. The amino acid sequences of active mutants revealed a strong preference for the replacement of Met213 with a His residue. This substitution also had the greatest observed effect on specificity, conferring a greatly increased and, in some cases, dominant ability to cleave at His residues in synthetic amide substrates. Mutant enzymes with greatly increased proteolytic activity were also found in the library.

The aim of enzyme engineering is to generate proteins with new and useful functional properties. The ability to alter the substrate specificity of an enzyme would be particularly advantageous, and many attempts have been made to change this property using site-directed mutagenesis. While knowledge-based engineering has been rewarding, its success has been limited by large gaps in the current understanding of protein folding and protein–ligand interactions (e.g., Craik et al. (1985), Wilks et al. (1988, 1990), Rutter et al. (1987), Henderson et al. (1991), and Alexander et al. (1991)). In a complementary approach, combinatorial random substitution has been used to generate libraries of variant proteins that contain a proportion of functional mutants (Reidhaar-Olson & Sauer, 1988; Lim & Sauer, 1989, 1991), these often having enzymatic properties different from those of the parent (Oliphant & Struhl, 1989; Dunn et al., 1988; Dunn & Jennings, 1992). Proteases constitute an industrially useful group of enzymes for which substrate specificity is of the utmost importance. While purely rational approaches have been used to change the substrate specificities of several proteases (e.g., Craik et al. (1985), Wells and Estell (1988), Beaumont et al. (1992), Wilson et al. (1991), Carter and Wells (1987), Khouri et al. (1991), and Hedstrom et al. (1992)), there are few examples of combinatorial random substitution being applied to this task (Evnin et al., 1990; Teplyakov et al., 1992). To determine whether proteases of novel specificity could be generated in this way, we simultaneously randomized several of the amino acid residues thought to influence the primary specificity of a serine protease.

The protease chosen for manipulation was α-lytic protease, a serine protease secreted by the soil bacterium *Lysobacter enzymogenes* (Whitaker, 1970). The structure of the protease, which has been determined to high resolution (Fujinaga et al., 1985), shows the enzyme to be a chymotrypsin homologue. Like elastase, α-lytic protease preferentially cleaves on the C-terminal side of small uncharged residues such as Ala (Kaplan et al., 1970; Bauer et al., 1981). The residues responsible for this primary cleavage specificity may be deduced from several crystallographic structures of enzyme–inhibitor complexes (Bone et al., 1989a, 1991a,b). Small amino acid residues such as Ala are preferred at the scissile bond because the pocket that accommodates the substrate P_1 residue is shallow,[1] largely due to the presence of two bulky Met residues at this subsite (Met190 and Met213, in the revised numbering system of Fujinaga et al. (1985)). Replacement of either of these with Ala residues results in a mutant enzyme that prefers large hydrophobic residues in the P_1 position (Bone et al., 1989b). Since we felt that an enlarged S_1 pocket allowed a greater scope for substitution at other positions contributing to this subsite, we selected one of these mutants (Met190Ala; see Figure 1) as the parent for a protease library.

Crystallographic structures reveal that the S_1 pocket of the Met190Ala mutant (Bone et al., 1991a) is largely defined by

† This work was sponsored by the Australian Government, Dept. of Industry, Technology & Commerce, in conjunction with Peptide Technology Ltd. and Burns Philp Ltd., under the Industry Research & Development Generic Biotechnology Program (Agreement No. 14026).

* Author to whom correspondence should be addressed.

‡ Present address: Chief, C.S.I.R.O. Division of Tropical Animal Production, Private Bag No. 3, P.O. Indooroopilly, Queensland 4068, Australia.

[1] In the nomenclature of Schecter and Berger (1967), the substrate residue immediately N-terminal to the scissile bond is termed the P_1 residue, the one before that is the P_2 residue, and so on. Residues C-terminal to the scissile bond are termed P_1', P_2', etc. Cognate residue-binding subsites in the protease are identified by use of the letter S in place of P, e.g., S_1', S_2', etc.

0006-2960/93/0432-6250$04.00/0 EXHIBIT G

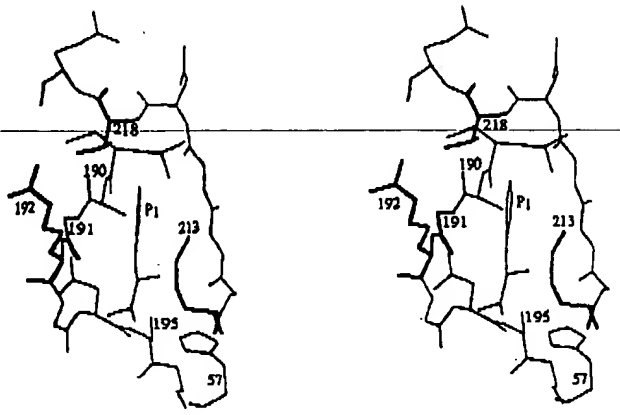

FIGURE 1: Stereoview of the active site of the Met190Ala mutant of α-lytic protease. This enzyme was used as the parent for the library described in this article. The viewer looks directly into the S_1 binding pocket, in which a substrate P_1 side chain (Phe) is shown; the four residues in the protein targeted for random substitution are shown by thick lines. This diagram is derived from file 1P08 in the Brookhaven Protein Data Bank (Abola et al., 1987).

two segments of polypeptide, which we labeled A and B:

segment A segment B

190 192 194 213 215 218 220 226

—Ala*Gly**Arg**Gly**Asp* — // —Met*Ser**Gly**Gly**Asn*Val —*Asn* —*Ser* —

Structural information and sequence homology with related enzymes indicate that the residues shown in boldface italics are essential to the overall structure of the protein and suggest that those shown in lightface italics may also be important. Position 190 and our reason for choosing Ala at this position have been discussed above. We did not consider any of the four remaining residues at the S_1 subsite (marked with asterisks) to be essential to the overall architecture of the protein. Our library was therefore constructed using targeted random mutagenesis to effect simultaneous random substitutions at some or all of the positions 191, 192, 213, and 218 (Figure 1).

EXPERIMENTAL PROCEDURES

Materials. The limited range of commercially available amino acid–chromophore conjugates meant that our suite of synthetic substrates could not all be made with the same leaving group. In consequence, screening was done using some chromogenic substrates of the form *Suc*AlaProXaa-pNa[2] and some fluorogenic substrates of the form *Suc*AlaProXaa-βNap (where Xaa denotes any natural amino acid residue or D-Ala). With the exception of *Suc*AlaProAla-pNa (Peptide Institute, Japan), these substrates were synthesized from (*tert*-butyloxycarbonyl)-AlaPro and the appropriate amino acid-pNa or -βNap, succinylated using succinic anhydride, and purified by reversed-phase HPLC. In each case, the composition was confirmed by amino acid analysis. Peptides with P_1 residues as follows were verified as enzyme substrates by digestion with the enzyme named: Ala, D-Ala, Gly, Leu, Met, and Val, wild-type α-lytic protease; Arg, bovine trypsin (Sigma); Asp and Glu, endoprotease Glu-C (Boehringer Mannheim); Phe, α-chymotrypsin (Sigma); and Pro, proline-specific endoprotease (Seikagaku Kogyo, Japan). Substrate purity was estimated in each case from the total amount of *p*-nitroaniline released. Any additional peptide substrates were purchased from Bachem. Molecular biology and related procedures were according to Sambrook et al. (1989), unless otherwise specified.

Cloning and Sequencing. We used PCR (95 °C, 30 s; 50 °C, 30 s; 72 °C, 1.5 min; 30 cycles) with primers having sequences 5′-ATTTATGCATGCCGATCAGGTCGATCCTCAG-3′ and 5′-TCTCATCGATCTATTAACCCGTGACCAGGCTCAGGCC-3′ in the presence of 7-deaza-2′-dGTP (Innis, 1990) to amplify the segment encoding the pro and enzymatic regions of the α-lytic protease gene (Silen et al., 1988) from *L. enzymogenes* genomic DNA. After it was end-filled with Klenow, the blunt-ended fragment was cloned into Bluescript (Stratagene Inc.). Dideoxy sequencing of DNA was done with Taq polymerase using end-labeled primers in the presence of 7-deaza-2′-dGTP; the best results were obtained when undenatured DNA (250 ng) was used as the template for the Promega "fmol" linear amplification method (95 °C, 30 s; 70 °C, 30 s; 30 cycles). After verification of the cloned sequence, a fragment encoding the *pelB* secretion leader (Lei et al., 1987) was inserted upstream of the protease gene, and the resulting cassette was ligated into pBS(+) (Stratagene Inc.) so that it would be expressed by the indigenous *lac* promoter (Silen et al., 1989; K. D. Haggett, unpublished results).

Mutagenesis. A *Bal*1–*Eco*R1 fragment (0.36 kb) that encoded the active-site region of the protease was subcloned into M13mp19, and the oligonucleotide 5′-CCGAATCGCCGCGGCCCGCGCAGGCGTTGC-3′ was used (Nakamaye & Eckstein, 1986; Sayers et al., 1988) to introduce the mutation Met190Ala. Further mutagenesis of the Met190Ala mutant was effected using the oligonucleotide 5′-TTGCCGTTGGACTGTGAGTTGCCGCTGCTCATCACGCCCTGC-3′ to introduce an *Eco*B restriction site (TGAN$_8$TGCT) into the region encoding the S_1 pocket, which resulted in two amino acid substitutions (Gly215Ser and Val218Ser). Replacement of the corresponding portion in the pBS(+) expression construct with the mutated *Bal*1–*Eco*R1 fragment gave us the construct to be used as the template for the library. The *Eco*B site served as a genetic marker and later allowed us to select against the unmutated template sequence (Carter, 1991).

Oligo-A, which had the sequence 5′-CACGAACCGCCCGAATCGCC^{G}/$_{C}$NN^{G}/$_{C}$NNCGCGCAGGCGTTGCCCTGG-3′, was designed to bind to the sense strand encoding segment A in the library template and to effect random substitutions at positions 191 and 192. Oligo-B, which had the sequence 5′-CAGTTGTTGCCGTTGGACTG^{G}/$_{C}$NNGTTGCCGCCGCT^{G}/$_{C}$NNCACGCCCTGCGCCTGGCCGG-3′, was designed to bind to the sense strand encoding segment B in the library template and to effect random substitutions at positions 213 and 218, as well as to restore Gly at position 215. The codon format NN^{G}/$_{C}$ ensured that each amino acid could be represented and eliminated two of the three stop codons (Dunn et al., 1988).

Mutagenic priming by oligo-B destroyed the *Eco*B selection sequence, allowing direct selection for incorporation of this oligonucleotide. To effect targeted random mutagenesis, existing methods (Foss & McClain, 1987; Kramer & Fritz, 1987; Inouye & Inouye, 1991) were adapted to maximize coupled priming. A sample of double-stranded library template construct was digested with *Eco*N1 and *Eco*R1 to remove a small region (0.2 kb) around the target sequence (i.e., the DNA encoding segments A and B), while another sample was cleaved at one position only using *Hin*dIII. About 460 ng of the *Eco*N1–*Eco*R1 fragment (4.3 kb) and 50 ng of the *Hin*dIII fragment were denatured together (100 °C, 3

[2] Abbreviations: *Suc*, succinyl; pNa, *p*-nitroanilide; βNap, β-naphthylamide.

min) and reannealed (65 °C, 10 min and then cooled to 0 °C over 30 min) in 10 μL of 100 mM Tris containing 500 mM KCl, pH 8, to generate approximately 50 ng of gapped duplex. At this stage, 8.5 pmol of phosphorylated oligo-A and 8.5 pmol of phosphorylated oligo-B were added, and the mixture was incubated at 65 °C for 3 min and then cooled to 0 °C over 30 min. T4 DNA polymerase (0.17 unit) and *E. coli* DNA ligase (0.85 unit) were added, and extension/ligation was achieved by incubation in 22.4 μL of 50 mM Tris containing 18 μg/mL T4 Gene 32 protein, 0.19 mM NAD^+, 0.23 mM each dNTP, 4.7 mM dithiothreitol, 35 mM KCl, 56 mM ammonium acetate, and 4.7 mM $MgCl_2$, pH 8, at 28 °C for 3.3 h. The DNA was ethanol precipitated and electroporated into *E. coli* HB2155 (Carter et al., 1985) using a BioRad gene pulser (five 40-μL aliquots at 2.5 kV, 25 μF, 600 Ω, with an electrode gap of 0.2 cm). Constructs retaining the *EcoB* site (i.e., those that had failed to incorporate oligo-B) were destroyed in this *E. coli B* strain. Following discharge, cells were allowed to recover in 1 mL of SOC broth (Sambrook et al., 1989) for 50 min at 28 °C before the titer of primary transformants was measured. The SOC outgrowths were then pooled and used as the inoculum for 50 mL of L-broth containing 0.2 mg/mL ampicillin and 2% (w/v) glucose, which was agitated at 25 °C for 21 h (A_{660} = 3.9). Plasmid isolation using a Quiagen-100 column (Diagen GmbH) yielded 67.5 μg of library DNA. Electroporation of *E. coli* JM109 with a small portion (100 ng) of this provided a sample of transformants in an amber suppressor strain, where the only nonsense codon possible at randomized triplets (TAG) would be suppressed.

Screening Methods. Preliminary screening of transformants was done by culturing colonies at 25 °C on L-agar plates (pH 7.2) containing 2% skim milk powder, where secretion of active protease resulted in localized clearing of the opaque growth medium due to degradation of the casein. After 11 days, the ratio of the diameter of the cleared zone to that of the colony was converted to an index value on a scale of 0–20. Obviously, this screening procedure (and those described below) can only identify enzymes that retain activity long enough after secretion to exert a detectable effect. In consequence, there is a bias toward active enzymes that also possess reasonable stability.

The substrate preferences of active proteases were first characterized in a qualitative fashion by a substrate overlay method, as follows. Colonies were grown in separate compartments of multiwell plates (such as Nunclon Δ Multidishes) on LAmp-broth (pH 7.5, 50 μg/mL ampicillin) that had been solidified with 1.5% (w/v) Seaplaque (FMC Marine Colloids). Growth for 10 days at 25 °C produced sufficient enzyme for sensitive detection. At this time, each well was filled with 1 mL of a molten solution (40 °C) containing 2 mM pNa or 10 mM βNap substrate, 1% (w/v) Seaplaque, and 11% *N,N*-dimethylformamide in 100 mM Hepes buffer, pH 8.0. Plates were incubated at 37 °C and inspected repeatedly over a period of 48 h to estimate visually the release of yellow *p*-nitroaniline or fluorescent β-naphthylamine (using illumination at 366 nm for the latter). Finally, to enhance detection of low levels of *p*-nitroaniline or β-naphthylamine, plates were developed by diazotization (Ohlsson et al., 1986) or by reaction with Fast Blue Salt BN (Barrett, 1972), respectively. Transformants were scored on a nonlinear scale of 0–20 to reflect the observed rate of substrate hydrolysis. The partly logarithmic nature of the scale emphasizes weak activities.

Initial rates of substrate hydrolysis in solution were measured spectrophotometrically at 410 nm for pNa (Erlanger et al., 1961) and at 340 nm for βNap substrates (Lee et al., 1971). Because reaction rates only afford a true reflection of k_{cat}/K_m when the substrate concentration is small relative to K_m, we used the lowest substrate concentrations consistent with workable assay durations. Reactions were conducted at 25 °C in 100 mM Hepes, pH 8, containing 9% *N,N*-dimethylformamide and pNa substrate (0.5 mM) or βNap substrate (5 mM). Under these conditions, the amount of enzyme activity releasing 1 μmol of chromophore/fluorophore per minute was defined as 1 unit.

RESULTS

Construction of the Library. The gene in the expression construct used as the template for the library encoded the Met190Ala variant of α-lytic protease and also contained a selectable genetic marker that directed the additional substitutions Gly215Ser and Val218Ser. When the template construct was cultured in *E. coli* JM109 on skim milk plates, the triple-mutant enzyme it produced (Met190Ala/Gly215Ser/Val218Ser) was shown to be inactive. Molecular modeling (not shown) had indicated that this would be the case, since the side chain of Ser215 protrudes into the space normally occupied by the main-chain atoms of the substrate P_2 residue. It was advantageous that unmutated template constructs escaping the genetic selection did not give rise to active enzymes in the library. The targeted random mutagenesis procedure was designed to remove the inactivating mutation at position 215 and to randomize up to four other positions (namely, 191, 192, 213, and 218) at the S_1 subsite. Combinatorial replacement by all 20 amino acids at each of the target positions permits a total of 1.6×10^5 permutations. The mutagenesis procedure generated a library containing 9.7×10^4 primary transformants.

Active Enzymes in the Library. Preliminary tests with endoprotease Glu-C, endoprotease Lys-C, and trypsin confirmed that even proteases with narrow substrate specificities could be detected using the skim milk screen. When 8.4×10^3 clones from the library were cultured in *E. coli* JM109 on skim milk plates, 0.57% of the colonies expressed active enzymes. All 47 of the active enzymes in this sample of the library hydrolyzed casein with activities equal to or greater than that of the parent (Table I), and a number (such as mutants 1 and 2) were exceptionally active in this assay.

Using wild-type and Met190Ala protease (WT and MA, Table I), the relative magnitudes of scores from the plate-overlay screening procedure were shown to be in general agreement with those of reaction rates determined spectrophotometrically.[3] Rate values for βNap substrates had to be adjusted before inclusion to compensate for the different (and usually slower) rate of hydrolysis of the leaving group, whereas this difference was accommodated in substrate overlays by using different scoring systems for βNap and pNa substrates. Reaction rates at the chosen substrate concentrations were shown to provide a reliable reflection of the relative magnitudes of specificity constants (k_{cat}/K_m) published for the test enzymes (Wilson et al., 1991; Bone et al., 1989b).

[3] An apparent discrepancy arises at the upper limit to the scale for plate screen scores, where very strong activities are all awarded the maximum value of 20 points. While discrimination between strong reactions is not possible in the plate screen due to rapid color or fluorescent saturation of the culture wells, such restrictions do not apply to the spectrophotometric rates. Thus with mutant 55 it is possible for spectrophotometric rates as varied as 1437 units/L (P_1 = Leu), 4207 units/L (P_1 = Ala), 10 741 units/L (P_1 = Phe), and 13 674 units/L (P_1 = Met) all to be awarded the same (maximum) score in the screen data.

Table I: Activity Data for Reference Enzymes and Enzymes from the Library

			pNa substrates, P_1 as shown[a]										βNap substrates, P_1 as shown[a]									positions[a]					
no.[b]	data[c]	SM[d]	G	dA	A	V	L	D	E	M	F	R	S	T	I	N	Q	H	Y	W	Adj[e]	191	192	213	218	n[f]	%[g]
Reference Enzymes[h]																											
WT	Sc	11	2	+	20	6	0	0	0	2	0	0	+	0	0	0	0	0	0	0		G	R	M	V		
	Ra		+	0	140	3	0	0	0	4	0	0	2	0	0	0	0	0	0	0	5.8						nd
	Pu		33		2100	79	+			180	+		200	20	+	3	68										
MA	Sc	3	+	0	6	2	12	0	0	20	20	0	0	1	0	+	0	0	2	0		G	R	M	V	0	
	Ra		0	0	10	1	67	0	0	308	321	0	+	+	0	+	0	0	13	0	5.0						60
	Pu				1000	300	11000			35000	31000																
Active Library Enzymes[h,i]																											
1	Sc	18	3	0	3	0	+	0	0	20	10	0	+	0	0	8	1	16	0	0		G	V	H	M	1	
	Ra		2	0	4	0	0	+	0	87	8	0	0	2	0	7	2	92	+	+	3.7						nd
2	Sc	20	3	0	3	0	0	1	0	16	10	0	+	0	0	8	1	16	0	0		G	R	H	M	1	
7	Sc	18	2	0	2	0	1	+	0	12	2	0	0	0	0	2	0	12	0	0		G	R	H	I	1	
	Ra		+	0	2	0	0	0	0	55	3	0	+	0	0	6	+	117	0	+	2.6						60
9	Sc	13	2	0	6	0	+	+	0	12	6	0	0	0	0	1	0	12	0	0		G	R	H	L	4	
	Ra		1	0	5	0	0	0	0	37	6	0	+	0	0	3	+	196	0	0	4.3						95
10	Sc	9	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	+	0	0		G	R	H	E	1	
11	Sc	10	+	0	+	0	0	+	0	1	+	0	0	0	0	0	0	2	0	0		G	T	H	S[j]	2	
	Ra		0	0	+	0	0	0	0	1	0	0	+	0	0	+	0	+	+	+	0.5						nd
12	Sc	4	0	0	0	0	0	1	0	3	1	0	0	0	0	0	0	4	0	0		G	R	H	S	2	
14	Sc	8	1	0	1	0	0	1	+	6	2	0	0	0	0	+	0	2	0	0		G	R	H	A	2	
	Ra		+	0	+	0	0	0	0	7	+	0	0	0	0	+	+	4	0	0	nk						nd
16	Sc	10	0	0	+	0	+	0	0	3	2	0	0	0	0	0	0	4	0	0		S	M	H	L	1	
19	Sc	10	+	0	+	0	0	+	0	1	+	0	0	0	0	+	0	1	0	0		G	H	H	T	1	
	Ra		0	0	0	0	0	0	0	+	0	0	0	0	0	+	0	+	0	0	nk						nd
22	Sc	14	2	0	1	0	+	+	0	12	2	0	0	0	0	1	0	4	0	0		G	R	H	V	3	
	Ra		+	0	+	0	0	0	0	20	1	0	0	0	0	1	+	22	0	+	1.5						st
36	Sc	7	+	0	1	0	0	+	0	1	+	0	0	0	0	0	0	2	0	0		G	R	H	N	4	
	Ra		0	0	+	0	0	0	0	+	+	0	+	0	0	0	0	5	0	0	1.1						nd
17	Sc	4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	+	0	0		Q	A	I	T[k]	1	
31	Sc	9	3	0	16	10	16	0	0	20	20	0	4	1	0	1	0	1	12	0		G	R	M	L	7	
35	Sc	15	3	0	16	2	16	0	0	20	20	0	2	1	0	+	0	+	8	0		G	L	L	M	1	
37	Sc	15	2	0	16	3	16	0	0	20	20	+	2	1	0	+	0	+	12	4		A	Q	T	M	1	
38	Sc	11	0	0	3	2	16	0	0	20	20	0	0	+	0	+	0	+	2	1		G	S	T	V	1	
39	Sc	8	2	0	20	+	12	0	0	16	20	0	+	+	0	+	0	0	4	+		G	R	F	M	1	
	Ra		2	0	198	0	26	0	0	1443	1238	0	2	+	0	2	0	2	94	2	8.2						st
41	Sc	14	6	+	20	12	20	+	+	20	20	+	2	1	+	2	1	1	16	4		G	R	T	M	1	
	Ra		5	0	1071	11	630	0	+	6646	6514	+	2	1	+	7	1	10	1013	52	4.8						50
43	Sc	7	1	0	10	10	16	0	0	20	20	0	2	1	0	1	0	+	8	0		G	R	M	I	2	
44	Sc	8	1	0	10	10	20	0	0	20	20	0	0	1	0	0	0	+	2	1		A	T	T	I	1	
45	Sc	7	2	0	20	1	12	0	0	20	20	0	1	1	0	0	0	0	1	+		A	W	T	L	1	
	Ra		+	0	305	+	24	0	0	228	133	0	2	0	0	0	0	0	6	1	5.0						15
47	Sc	9	2	0	20	12	12	0	+	20	20	0	2	1	1	1	2	2	4	1		G	R	S	L	3	
	Ra		+	+	148	25	50	+	+	583	546	0	7	1	3	3	5	28	70	4	6.0						20
50	Sc	15	1	0	16	10	16	0	0	20	20	0	2	1	0	+	0	1	12	0		G	R	M	M	1	
51	Sc	9	3	0	16	1	16	0	0	20	20	0	2	1	0	+	0	1	8	+		G	R	L	L	2	
55	Sc	11	+	+	20	12	20	+	+	20	20	0	2	1	0	2	1	2	16	4		G	R	T	L	1	
	Ra		37	0	4207	23	1437	+	2	13674	10741	0	8	+	0	8	+	6	480	16	2.4						15

[a] Substrate P_1 residues and amino acid residues at target positions in the enzymes are indicated using one-letter codes; dA is D-Ala. Since no enzyme in this table had detectable activity with substrate *Suc*AlaProPro-pNa, there is no column for P_1 = Pro. [b] Enzyme identifier: WT, wild-type α-lytic protease; MA, Met190Ala mutant (library parent); numbers, library enzymes. [c] Sc: Plate screen scores (substrate overlay) on pseudo-log scale. Ra: Spectrophotometric rates (units/L). Pu: One-tenth of published values for k_{cat}/K_m (s^{-1} M^{-1}) with *Suc*AlaAlaProXaa-pNa (where Xaa denotes an amino acid residue) taken from Wilson et al. (1991) and Bone et al. (1989b). [d] SM: Index of skim milk clearing (see text). [e] Adj: The adjustment factor applied to rate data to normalize for different leaving groups (see text). nk: Factor not known (no normalization applied). [f] n: Number of instances of each sequence in the 47 active library enzymes sequenced. [g] %: The approximate percentage of activity remaining in cell-free supernatant after 260 days at 4 °C, pH 9. nd: not determined. st: stable over at least 80 days. [h] +, value below 1 (number omitted to improve clarity). [i] Active library enzymes are segregated into those with His213 (top set) and those without (bottom set). [j] Spurious mutation Ser219aAla also present. [k] Spurious mutations Gln219His and ΔSer219a also present.

With this information in hand, all 47 of the active enzymes in the library sample were screened by substrate overlay (Table I). In addition, 13 transformants that expressed active enzymes were cultured in rich liquid medium. Since all of the liquid cultures contained comparable levels of mutant enzyme protein (as judged by HPLC), and zymograms (not shown) confirmed the absence of other proteases, samples of cell-free supernatants were used directly in spectrophotometric assays to measure reaction rates. The scaling factors used to adjust the rates of βNap substrates (calculated by comparing the activities of each mutant on the *Suc*AlaProAla-pNa and *Suc*AlaProAla-βNap substrates) are shown in Table I. As before, spectrophotometric rates and screening scores were in broad agreement. Assays that monitored the hydrolysis of ester substrates (benzyloxycarbonyl)-AlaLys-OMe and (benzyloxycarbonyl)-AlaCys-OMe by measurement of proton release (R. G. Whittaker, unpublished results) failed to find any library enzymes (among the 13) with a major capacity for cleavage at the two natural P_1 residues not already tested (data not shown). Moreover, spectrophotometric measurement of the rate of hydrolysis of tetrapeptide *Suc*AlaAla-ProLys-pNa by the Met190Ala enzyme and library enzymes 7, 9, and 55 confirmed that none had a major ability to cleave this substrate (data not shown).

From Table I it is evident that about one-half of the enzymes in the library sample exhibited substrate preferences that

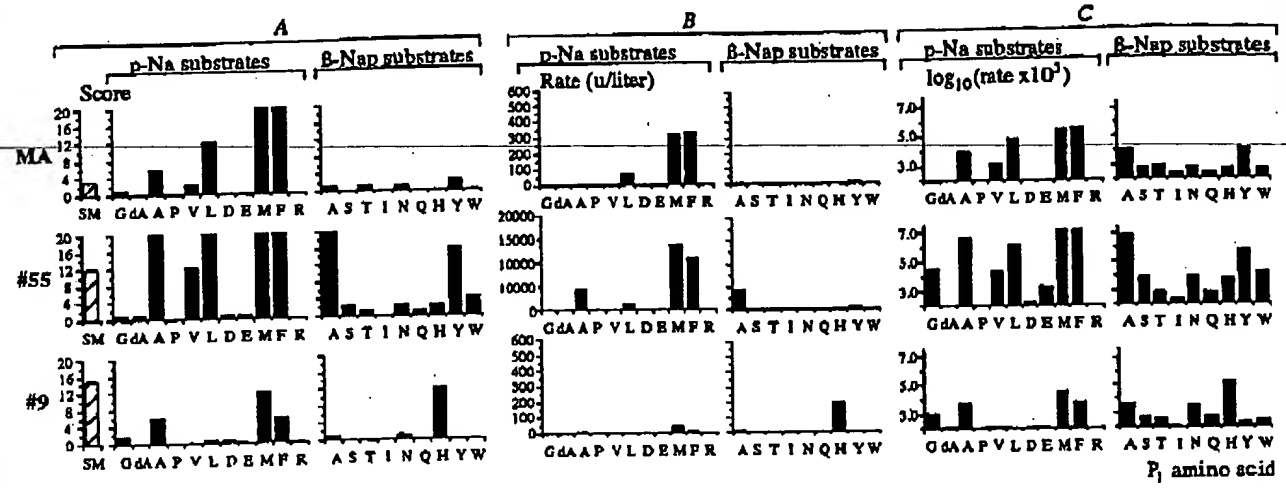

FIGURE 2: Activity data for the library parent (MA) and two enzymes found in the library (mutants 55 and 9). Three histograms are shown for each enzyme: (A) the plate-screen scores for skim milk clearing (hatched bars) and substrate overlays (solid bars); (B) linear plot of spectrophotometric rate data (expressed as units of activity per liter of supernatant); (C) $\log_{10}$ plot of the spectrophotometric rate data. In addition to skim milk clearing (SM), the horizontal axes show substrate P_1 residues in one-letter codes; dA is D-Ala. Vertical axes without markings relate to the first marked scale to the left. The partly logarithmic nature of the plate-screen scores (explained in Experimental Procedures) is evident. Logarithmic scales serve to emphasize weak activities.

reflected the specificity of the parent protease, giving the best scores with the Met and Phe substrates and good scores with Ala and Leu. The remaining enzymes constituted a distinct group that showed greater selectivity and preferred to cleave the His and Met substrates. In the enzymes of both groups, additional (but lesser) cleavage capabilities were featured to different extents. There were also large differences between the overall levels of activity exhibited by individual enzymes, with some (e.g., mutants 41 and 55) showing rates 20–45 times faster than that of the library parent or wild-type protease. Some activity data for selected enzymes have been plotted to illustrate the functional diversity contained in the library (Figure 2). In general, however, we found that enzymes in the library had little or no ability to act on charged P_1 residues (His being predominantly uncharged at the pH of the assay) or to accommodate those which were sterically unusual (such as D-Ala and Pro).

Substitutions in the Library. DNA sequencing of 21 clones selected at random from the library (not shown) revealed that 29% of the transformants were invalid, mostly because of partial incorporation of oligo-B (resulting in destruction of the *EcoB* selectable marker but retention of the inactivating mutation Gly215Ser). Some 52% of the transformants were valid clones that resulted from coupled priming by oligo-A and oligo-B (changes possible to positions 191, 192, 213, and 218), while the remaining 19% of the transformants were valid clones that resulted from mutagenic priming by oligo-B alone (changes possible to positions 213 and 218 only). Sequence data (not shown) revealed that the incorporation of nucleotides at randomized positions was somewhat biased, with different nucleotides being favored in different positions in an unpredictable fashion (cf. Dunn et al. (1988) and Oliphant and Struhl (1989)). Nevertheless, the extent of nucleotide substitution was sufficient for adequate diversity to occur at the amino acid level (see below), and the biases cannot have been seriously limiting because many clones encoding active enzymes featured substitutions poorly represented in the library. We presume that the biases inherent to the library served mostly to reduce its yield of active mutants. As statistically expected from the frequency of active enzymes in the library (0.57%), all 21 clones picked at random encoded inactive enzymes. Amino acid combinations for the valid members of this sample are shown in Table II.

Table II: Inactive Combinations[a]

	position			
	191	192	213	218
103	Pro	Arg	Thr	Arg
104	Val	Lys	Met	His
107	Gly	Arg	Met	Cys
108	Tyr	Lys	His	Ala
109	Glu	Gly	Met	Leu
110	Val	Asn	Gln	Ala
112	Gly	Arg	Arg	Gln
113	Tyr	Lys	Leu	Arg
114	Gly	Arg	Pro	Ser
121	Gly	Arg	Asn	Ser
124	His	Asn	Lys	Asn
128	Gly	Tyr	Arg	Ala
129	Gln	Cys	Leu	Lys
146	Lys	Val	Pro	Thr

[a] Column 1 contains the clone identification numbers. Each combination was observed only once.

The substitutions present in each of the 47 active enzymes from the library sample (see above) are listed in Table I. About 75% of the active enzymes resulted from mutagenic priming by oligo-B alone (changes possible to positions 213 and 218 only). The remaining active mutants resulted from coupled priming by oligo-A and oligo-B (changes possible to positions 191, 192, 213, and 218). Many residue combinations were observed more than once (Table I) to the extent that about one-half of the active enzymes were sequence replicates. Only slight differences (not shown) were observed in the substrate preference profiles of mutants having the same sequence. We thought it useful to have independent clones of particular mutant genes, since the existence of replicates would to some extent offset the effects of any spurious mutations in library clones. However, one of the few unplanned changes detected (Ser219aAla) was found to occur in two clones having the same S_1 sequence (mutants 8 and 11). Interestingly, the only other spurious mutation observed (Gln219His/ΔSer219a) affected the same position in the protein, which was located in a surface loop adjacent to the S_1 site. While changes in this region are thought to be unlikely to influence enzymatic activity, we have been careful to draw

our main conclusions about structure–activity relationships from trends rather than from data for individual mutants.

Structure–Activity Relationships. Figure 3 compares the frequency of substitutions observed in the active enzymes with data for substitutions in the (inactive) clones chosen at random from the library. While there was some bias in the distribution of residues in the randomly chosen clones, representation of the 20 alternatives was broadly comparable. In contrast, selection for biological function of the mutants resulted in substitutions being confined to more limited sets of amino acids. In active enzymes, Pro (which might be disruptive to structure) was avoided at all four of the positions open to substitution. Given that His should be largely unprotonated, charged residues were scarce. The lack of charged substituents was surprising, considering that all four positions have access to solvent and than one of them (position 192) is occupied by Arg in the wild-type and parental sequences. Cys was also avoided in active enzymes, perhaps because disulfide bond formation in the protein is vulnerable to interference from additional Cys residues. In this connection, we note that mutants 31, 43, and 50 (identical but for the substitution of Leu, Ile, and Met, respectively, at position 218) were strongly active, whereas mutant 107 (identical to the previous mutants but for the substitution of Cys at position 218) was completely inactive (Tables I and II).

In addition to the general trends described above, each of the four randomized positions in active enzymes displayed characteristic substitution preferences. Position 191 showed a very strong preference for Gly (as found in the wild-type sequence) and accommodated other small residues with a frequency inversely related to their size, which suggested that this position was subject to tight steric constraints. Position 192 was quite permissive, allowing polar and nonpolar substitutions of very different sizes. Position 213 accommodated a limited set of residues, with a strong preference for His. Remarkably, all of the enzymes in the substrate-selective group that preferred to cleave His and Met substrates (Table I) were found to contain His at position 213. The remainder of the active mutants, which had broad substrate specificity, all contained residues other than His at this position, with Met (the wild-type residue) occurring most frequently. Position 218 accepted a range of residues, but had a strong preference for Leu. A more detailed examination (Figure 3b) revealed that the strong preference for Leu at position 218 was a feature of enzymes that did not contain a His at position 213. For His213 mutants, the preference for Leu at position 218 was slight, and there was a notable increase in the occurrence of small polar residues such as Ser and Asn.

Spectrophotometric rate data for the activity of mutant enzymes on substrates with different P_2 (and, in some cases, P_3 and P_4) residues are presented in Figure 4. The correlations presented in this figure indicate that different mutants responded similarly to the same change in the P_2 residue (and, where tested, in the P_3 and P_4 residues). This strongly suggests that subsites S_2–S_4 have not been greatly affected by the amino acid substitutions we introduced in the S_1 subsite. In contrast, comparison of reaction rates obtained for the action of each enzyme on *Suc*AlaProAla-pNa and *Suc*AlaProAla-βNap substrates (Adj, Table I) revealed differences in how mutants (other than those of identical sequence) responded to the switch in chromophore. This might be an indication of changes to the S_1' subsite caused unintentionally by our substitutions in the S_1 pocket. However, enzyme–substrate interactions C-terminal to the scissile bond are considered to be of limited importance in α-lytic protease (Bauer et al., 1981).

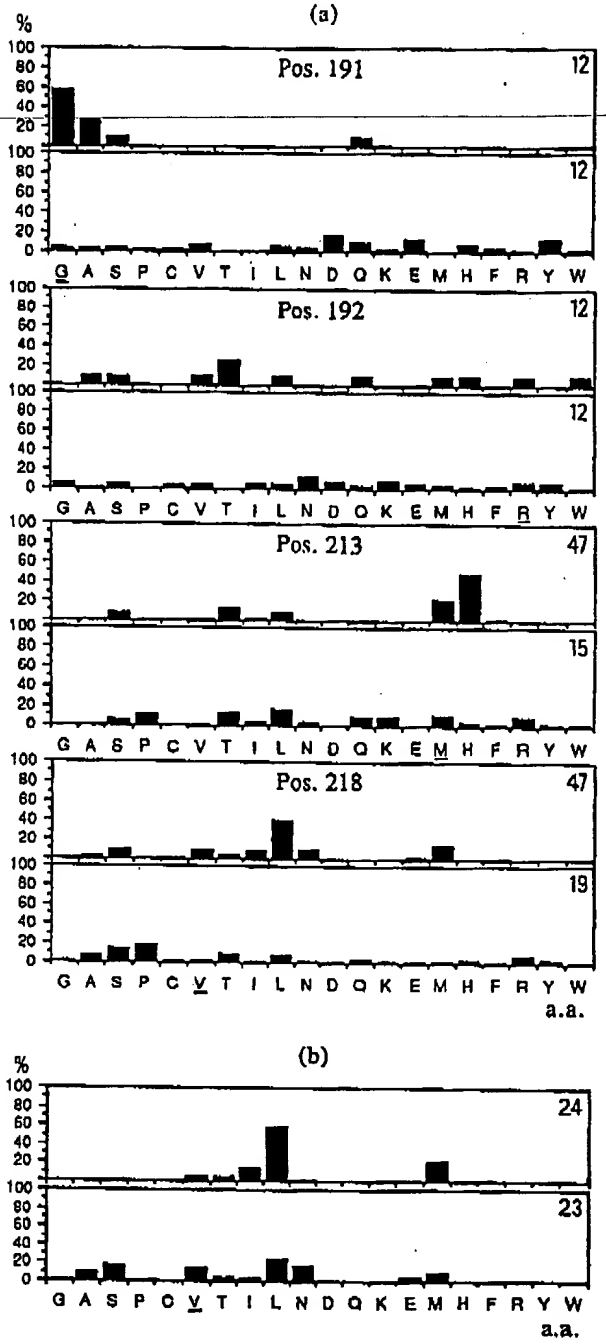

FIGURE 3: Frequency of substitutions in the library. (a) Each double panel refers to one of the target positions (Pos) in the protein and shows the incidence (%) of amino acid substituents (a.a.) in active library enzymes (upper panel), along with data for clones chosen at random from the library (lower panel). Amino acids are indicated using one-letter codes, and the wild-type residue for each position is underlined. The number in the top right-hand corner of each panel indicates the sample size (i.e., number of clones). Each upper panel is a histogram of the actual amino acid frequencies observed in active enzymes, excluding for positions 191 and 192 data from enzymes that were unmutated in segment A. In view of the more limited sample size for clones picked at random, each lower panel shows an amino acid distribution calculated from the observed nucleotide incorporation data for the corresponding codon. The histograms of actual amino acid substitutions (not shown) reflect these plots. (b) This double panel uses a similar format to show the incidence of amino acid substitutions at position 218 in active enzymes with residues other than His at position 213 (upper panel), along with the corresponding incidence in those having His at position 213 (lower panel).

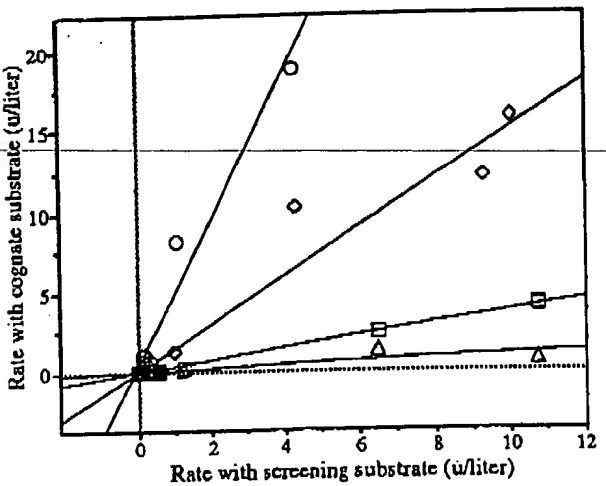

FIGURE 4: Effect of S_1 mutagenesis on S_2–S_4 binding sites. Using 12 library enzymes of different sequences, rates (expressed as units of activity per liter of supernatant) were measured with four pNa substrates, and each result was plotted against the rate obtained for the same enzyme with the same concentration of the corresponding screening substrate (0.5 mM). Cognate substrates were *Suc*AlaAlaValAla-pNa (○), *Suc*AlaAlaVal-pNa (◇), *Suc*AlaAlaPhe-pNa (□), and *Suc*GlyGlyPhe-pNa (△) for screening substrates *Suc*AlaProAla-pNa, *Suc*AlaProVal-pNa, *Suc*AlaProPhe-pNa, and *Suc*AlaProPhe-pNa, respectively. In order to fit all of the data on the same graph, rate values for the following substrates have been multiplied by the factors shown: *Suc*GlyGlyPhe-pNa, 50; *Suc*AlaAlaVal-pNa, 3000; *Suc*AlaProVal-pNa, 400.

A number of cell-free supernatants showed a steady decline in activity during the period of study (Table I). One of the most unstable enzymes (mutant 55) has broad specificity and very high activity, which suggests that autolysis may play a role in this process.

DISCUSSION

We have demonstrated that combinatorial random substitution of a few chosen positions in the substrate-binding site of a protease, followed by screening with multiple substrates, is a useful method for production and identification of mutant enzymes of altered structure and function. This approach is not limited to proteases and can be applied to any enzyme for which a structural model is available and for which appropriate assays can be devised. In fact, the approach we have adopted should be applicable in an iterative manner to enzymes for which structural information is limited.[4]

In our experiment, 0.57% of the library constructs expressed active enzymes, indicating the presence of some 550 active mutants in total. However, DNA sequencing revealed that 29% of the primary transformants were invalid (because of retention of Gly215Ser or for other reasons), indicating that the true frequency of active mutants must be about 0.8%. Moreover, sequencing also disclosed a level of duplication among clones that suggested the number of different active mutants in our library was closer to 300. Although most of the valid mutants in the library resulted from coupled priming by oligo-A and oligo-B, such double mutants (with four positions affected) retained activity much less often than mutants resulting from oligo-B alone (with only two positions affected). In consequence, while all of the active mutants in the library contained substitutions in the positions making the greater contribution to the S_1 pocket (namely, 213 and 218; Figure 1), only a limited number contained additional changes at the other two positions. This was a favorable situation for analyzing structure–activity relationships in the S_1 subsite.

Calculations indicated that 11.9% of the oligo-A sequences and 2.3% of the oligo-B sequences were compatible with enzymatic function.[5] This result strongly suggests that (when considered as pairs) positions 191 and 192 are much more tolerant of substitution than positions 213 and 218, an observation that is not directly apparent from independent consideration of substitutional preferences at each of these sites (Figure 3). The difference could be a reflection of the fact that positions 213 and 218 contribute considerably more to the surface that defines the S_1 pocket (Figure 1). Overall, the positions we investigated in α-lytic protease display a tolerance to substitution slightly greater than that calculated elsewhere for substrate-binding residues in β-lactamase (Palzkill & Botstein, 1992).

Because self-processing of the precursor at a Thr–Ala junction is required for release of the active mature protease (Silen et al., 1989), we expected that all of the active enzymes in our library would have a significant ability to cleave at Thr residues. However, we found that the activity of the wild-type protease and the library parent on a synthetic substrate with Thr as the P_1 residue was slight and that some of the enzymes in the library were even less active with this substrate. It seems that processing of the precursor junction by these enzymes is still sufficiently fast for good expression of the mature protease, perhaps due to the unimolecular nature of the reaction or due to unusual strain at this bond. Another surprise was the poor correlation between the activity of library enzymes on skim milk and on synthetic substrates ($r^2 = 0.8$ at best; data not shown). Transformants with the greatest capacity for milk clearing during growth (mutants 1 and 2) were not those with the highest overall activity when overlaid with synthetic substrates (mutants 41 and 55). Since the substrate preferences of the former group were not directed toward amino acids common in casein, and since we have evidence that our mutagenesis has not affected interactions at the other subsites critical to substrate binding (Figure 4), we are unable to offer a satisfying explanation for this phenomenon.

In general, residue changes in the protease that were compatible with enzymatic activity had only modest effects on substrate specificity (Table I). The one exception was the substitution of Met213 with its strongly favored replacement, His. This substitution severely reduced the (otherwise large) capacity of the enzyme for cleavage at Phe and conferred instead a greatly increased ability to cleave the fluorogenic substrate having His in the P_1 position. Neither the wild-type α-lytic protease nor the library parent (Met190Ala) displayed

[4] If an initial library is made by the randomization of a substantial number of different positions that may be involved in substrate binding, then a very low yield of active mutants should result. The sequences of these mutants will reveal which positions are largely restricted to the wild-type residue (such as position 191 in our study). A subsequent library in which such positions are fixed as the wild-type residue, while the remaining positions of the original choice are once again varied, should be much more productive in terms of active mutant enzymes and, for that reason, afford useful functional diversity.

[5] Oligo-B: Since 75% of the active enzymes (i.e., 75% of 0.57% of clones) resulted from priming by oligo-B alone, and 19% of the library constructs were valid clones that resulted from mutagenic priming by oligo-B alone, we concluded that 2.3% of the oligo-B sequences were compatible with enzymatic function. Oligo-A: Since 25% of the active mutants (i.e., 25% of 0.57% of clones) resulted from coupled priming by oligo-A and oligo-B, but only 2.3% of oligo-B sequences would be expected to allow activity, and 52% of the library constructs were valid clones that resulted from coupled priming, we calculated that 11.9% of the oligo-A sequences were compatible with enzymatic function.

any significant capacity for cleavage at His, whereas His was actually the preferred P_1 residue for some of the Met213His mutants (e.g., mutant 9). Thus it appears that the presence of a key residue in a key position can cause an abrupt switch in primary specificity.

In the absence of perturbations of pK_a in the S_1–P_1 complex, neither of the two imidazole rings should have been substantially ionized at the pH used in the study. We therefore chose to view His as an uncharged residue. The poor activity of library enzymes with substrates containing charged P_1 residues is likely to be a reflection of the fact that charged residues were almost absent from the repertoire of acceptable S_1 substituents, which would preclude the opportunity for electrostatic balance in enzyme–substrate complexes. Interestingly, a recently discovered homologue of α-lytic protease from *Streptomyces griseus* contains a His residue at position 213 (Svendsen et al., 1991) and cleaves preferentially at Glu in protein and synthetic substrates (Yoshida et al., 1988), even though molecular modeling predicts no charged amino acids near the S_1 region at the pH used. However, since the ability of the *S. griseus* protease (protease E) to cleave at His in protein or synthetic substrates is unknown, we are unable to compare it further with our Met213His enzymes. We do, however, note that *S. griseus* protease E contains a Ser at position 190. This invites speculation that a preference for His at position 213 is a response to the presence of a small residue at position 190 and would not have been observed if we had used wild-type α-lytic protease (rather than the Met190Ala variant) as the parent for our library.

In addition to their propensity for cleaving at His residues, all Met213His mutants efficiently cleaved the synthetic substrate with Met as the P_1 residue, further supporting the presumption that His213 is largely uncharged. Un-ionized His has polar aromatic character. Attention has recently been paid to the interactions of His with other aromatic residues in proteins (Lowenthal et al., 1992; Jasanoff et al., 1992), but His–His and His–Met interactions have not been studied in any detail. For our Met213His mutants of α-lytic protease, molecular modeling suggests that it is mainly the $N_{\epsilon 2}$–$C_{\epsilon 1}$ edge of the His213 imidazole that contacts the substrate P_1 residue, regardless of the precise orientation of the ring (L. D. Graham, unpublished results). In the more common histidine tautomer, i.e., the $N_{\epsilon 2}$-imino form (Walters & Allerhand, 1980), resonance theory (and Huckel molecular orbital calculations) predicts the $N_{\epsilon 2}$–$C_{\epsilon 1}$ edge to be electron-deficient. Modeling also suggests that His and Met are among the few uncharged P_1 side chains capable of positioning an electron-rich atom (the tertiary nitrogen and thioether sulfur, respectively) in good contact with this potentially electropositive region. The observed substrate preference of our Met213His enzymes may therefore be explained in terms of favorable polar interactions at the S_1 subsite.

We suggest from a comparison of mutants 7, 9, 14, 22, and 36 (identical but for the substitution of Ile, Leu, Ala, Val, and Asn, respectively, at position 218) that the overall activity of the His213 enzymes may increase in response to increased hydrophobicity of residue 218 (Table I). Inspection of the other rate data in Table I suggested that this effect may apply, at least loosely, to non-His213 enzymes as well. However, it was difficult to find further examples of this kind of structure–function correlation. While other residue substitutions in active enzymes resulted in substantial differences being observed in some minor activities (particularly in the ability to cleave at Val, Tyr, and Asn), changes in these activities did not show any systematic dependence upon the sizes or hydrophobicity indices of the S_1 substituents (not shown). We presume in such cases that many individual steric and electrostatic considerations combine to determine changes in specificity and that they do so in a nonadditive way that exceeds our present ability to model and understand them.

We do not consider that we have exhausted the functional diversity likely to be attained by this approach and expect that other mutants of altered primary specificity can be obtained by mutating a different combination of residues in the S_1 subsite, e.g., by randomizing position 190 in place of position 191. Moreover, we suggest that a library constructed to contain His at position 213 with random substitutions at other amenable S_1 positions (e.g., positions 190, 192, and 218) is likely to have a large proportion of His-cleaving proteases and may well include enzymes with tighter P_1 specificity for His. Work is in progress to test these hypotheses. In the meantime, we are engaged in the characterization of the more interesting of the existing library mutants in greater detail. The mutants most likely to find application as research enzymes (e.g., as tools in peptide mapping) are those with high activity and good selectivity, such as mutants 9 and 1. In addition, some library mutants (e.g., mutants 41 and 55) display a large increase in proteolyic activity over wild-type α-lytic protease and the library parent. Enzymes such as these, with very high activity and broad substrate specificity, could have applications as general-purpose reagents for protein degradation.

ACKNOWLEDGMENT

The authors are grateful to Mr. Denis Wright (C.S.I.R.O.) for suggestions and help in molecular biology, to Mrs. Gina Cherry (Peptide Technology) for technical assistance, and to Dr. Fred Widmer (Peptide Technology) for helpful discussions.

REFERENCES

Abola, E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases–Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., & Sievers, R., Eds.) pp 107–132, Data Commission of the International Union of Crystallography, Bonn.

Alexander, R. S., Nair, S. K., & Christianson, D. W. (1991) *Biochemistry 30*, 11064–11072.

Barrett, A. J. (1972) *Anal. Biochem. 47*, 280–293.

Bauer, C.-A., Brayer, G. D., Sielecki, A. R., & James, M. N. G. (1981) *Eur. J. Biochem. 120*, 289–294.

Beaumont, A., Barbe, B., Lemoual, H., Boileau, G., Crine, P., Fourniezaluski, M. C., & Roques, B. P. (1992) *J. Biol. Chem. 267*, 2138–2141.

Bone, R., Frank, D., Kettner, C. A., & Agard, D. A. (1989a) *Biochemistry 28*, 7600–7609.

Bone, R., Silen, J. L., & Agard, D. A. (1989b) *Nature 339*, 191–195.

Bone, R., Fujishige, A., Kettner, C. A., & Agard, D. A. (1991a) *Biochemistry 30*, 10388–10398.

Bone, R., Sampson, N. S., Bartlett, P. A., & Agard, D. A. (1991b) *Biochemistry 30*, 2263–2272.

Carter, P. (1991) in *Directed Mutagenesis: A Practical Approach* (McPherson, M. J., Ed.) pp 1–25, IRL Press, Oxford.

Carter, P., & Wells, J. A. (1987) *Science 237*, 394–399.

Carter, P., Bedouelle, H., & Winter, G. (1985) *Nucleic Acids Res. 13*, 4431–4443.

Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R., & Rutter, W. J. (1985) *Science 228*, 291–297.

Dunn, I. S., & Jennings, P. A. (1992) *Protein Eng. 5*, 441–446.

Dunn, I. S., Cowan, R., & Jennings, P. A. (1988) *Protein Eng. 2*, 283–291.

Erlanger, B. F., Korowsky, N., & Cohen, W. (1961) *Arch. Biochem. Biophys. 95*, 271–278.
Evnin, L. B., Vasquez, J. R., & Craik, C. S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6659–6663.
Foss, K., & McClain, W. H. (1987) *Gene 59*, 285–290.
Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1985) *J. Mol. Biol. 183*, 479–502.
Hedstrom, L., Szilagyi, L., & Rutter, W. J. (1992) *Science 255*, 1249–1253.
Henderson, G. B., Murgolo, N. J., Kuriyan, J., Osapay, K., Kominos, D., Berry, A., Scrutton, N. S., Hinchliffe, N. W., Perham, R. N., & Cerami, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8769–8773.
Innis, M. A. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J., Eds.) Academic Press, San Diego.
Inouye, S., & Inouye, M. (1991) in *Directed Mutagenesis: A Practical Approach* (McPherson, M. J., Ed.) pp 711–83, IRL Press, Oxford.
Jasanoff, A., Kochoyan, M., Fraenkel, E., Lee, J. P., & Weiss, M. A. (1992) *J. Mol. Biol. 225*, 1035–1047.
Kaplan, H., Symonds, V. B., Dugas, H., & Whitaker, D. R. (1970) *Can. J. Biochem. 48*, 649–658.
Khouri, H. E., Vernet, T., Menard, R., Parlati, F., Laflamme, P., Tessier, D. C., Goursalin, B., Thomas, D. Y., & Storer, A. C. (1991) *Biochemistry 30*, 8929–8936.
Kramer, W., & Fritz, H.-J. (1987) *Methods Enzymol. 154*, 350–367.
Lee, H.-J., La Rue, J. N., & Wilson, I. B. (1971) *Anal. Biochem. 41*, 397–401.
Lei, S.-P., Lin, H.-C., Wang, S.-S., Callaway, J., & Wilcox, G. (1987) *J. Bacteriol. 169*, 4379–4383.
Lim, W. A., & Sauer, R. T. (1989) *Nature 339*, 31–36.
Lim, W. A., & Sauer, R. T. (1991) *J. Mol. Biol. 219*, 359–376.
Lowenthal, R., Sancho, J., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 759–770.
Nakamaye, K., & Eckstein, F. (1986) *Nucleic Acids Res. 14*, 9679–9698.
Ohlsson, B. G., Westrom, B. R., & Karlsson, B. W. (1986) *Anal. Biochem. 152*, 239–244.
Oliphant, A. R., & Struhl, K. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9094–9098.
Palzkill, T., & Botstein, D. (1992) *Proteins: Struct., Funct., Genet. 14*, 29–44.
Reidhaar-Olson, J. F., & Sauer, R. T. (1988) *Science 241*, 53–57.
Rutter, W. J., Gardell, S. J., Roczniak, S., Hilvert, D., Sprang, S., Fletterick, R. J., & Craik, C. S. (1987) in *Protein Engineering* (Oxender, D. L., & Fox, C. F., Eds.) pp 257–267, Alan R. Liss, Inc., New York.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, CSH Laboratory Press, New York.
Sayers, J. R., Schmidt, W., & Eckstein, F. (1988) *Nucleic Acids Res. 16*, 791–802.
Schecter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Silen, J. L., McGrath, C. N., Smith, K. R., & Agard, D. A. (1988) *Gene 69*, 237–244.
Silen, J. L., Frank, D., Fujishige, A., Bone, R., & Agard, D. A. (1989) *J. Bacteriol. 171*, 1320–1325.
Svendsen, I., Jensen, M. R., & Breddam, K. (1991) *FEBS Lett. 292*, 165–167.
Teplyakov, A. V., van der Laan, J. M., Lammers, A. A., Kelders, H., Kalk, K. H., Misset, O., Mulleners, L. J. S. M., & Dijkstra, B. W. (1992) *Protein Eng. 5*, 413–420.
Walters, D. E., & Allerhand, A. (1980) *J. Biol. Chem. 255*, 6200–6204.
Wells, J. A., & Estell, D. A. (1988) *Trends. Biochem. Sci. 13*, 291–297.
Whitaker, D. R. (1970) *Methods Enzymol. 19*, 599–613.
Wilks, H. M., Hart, K. W., Feency, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., & Holbrook, J. J. (1988) *Science 242*, 1541–1544.
Wilks, H. M., Halsall, D. J., Atkinson, T., Chia, W. N., Clarke, A. R., & Holbrook, J. J. (1990) *Biochemistry 29*, 8587–8591.
Wilson, C., Mace, J. E., & Agard, D. A. (1991) *J. Mol. Biol. 220*, 495–506.
Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., & Makisumi, S. (1988) *J. Biochem. (Tokyo) 104*, 451–456.

JOURNAL OF BACTERIOLOGY, Feb. 1993, p. 858–865
0021-9193/93/030858-08$02.00/0

Vol. 175, No. 3

Combinatorial Mutagenesis of the *lamB* Gene: Residues 41 through 43, Which Are Conserved in *Escherichia coli* Outer Membrane Proteins, Are Informationally Important in Maltoporin Structure and Function

WAI CHI CHAN AND THOMAS FERENCI*

Department of Microbiology G08, University of Sydney, Sydney, New South Wales 2006, Australia

Received 11 August 1992/Accepted 11 November 1992

A new strategy for combinatorial mutagenesis was developed and applied to residues 40 through 60 of LamB protein (maltoporin), with the aim of identifying amino acids important for LamB structure and function. The strategy involved a template containing a stop codon in the target sequence and a pool of random degenerate oligonucleotides covering the region. In vitro mutagenesis followed by selection for function (Dex$^+$, ability to utilize dextrins) corrected the nonsense mutation and simultaneously forced incorporation of a random mutation(s) within the region. The relative importance of each residue within the target was indicated by the frequency and nature of neutral and deleterious mutations recovered at each position. Residues 41 through 43 in LamB accepted few neutral substitutions, whereas residues 55 through 57 were highly flexible in this regard. Consistent with this finding was that the majority of defective mutants were altered at residues 41 to 43. Characterization of these mutants indicated that the nature of residues 41 to 43 influenced the amount of stable protein in the outer membrane. These results, as well as the conserved nature of this stretch of residues among outer membrane proteins, suggest that residues 41 to 43 of LamB play an important role in the process of outer membrane localization.

Combinatorial mutagenesis studies are becoming important in the identification of important residues of proteins, including transport proteins (3, 15, 23, 25, 26). In this study, we developed a novel strategy for introducing random mutations into a defined region of maltoporin (LamB protein), with the aim of identifying residues significant in contributing to structure, assembly, and sugar transport. The strategy used with LamB can be applied to the mutagenesis of any gene coding for a selectable phenotype and has considerable advantages over cassette mutagenesis methods.

Maltoporin in the outer membrane of *Escherichia coli* has been favored as a model in studies of phage (lambda) binding (4) and sugar channel selectivity (1) as well as protein export and assembly (9, 20). Amino acid residues which play an important role in phage lambda and sugar binding have been identified (4, 14). However, the events leading to the localization and assembly of LamB trimers into the outer membrane are still not clear. In addition to the signal sequence, specific regions in the mature protein may also play an important role in the biogenesis process (9, 20) but remain to be convincingly identified.

There are several lines of evidence to suggest that residues 40 through 60 of mature LamB are of structural importance. Short in-frame deletions overlapping residues 39 to 49 of mature LamB resulted in the formation of unstable protein that was rapidly degraded (24), possibly as a result of incorrect outer membrane routing. Apart from its potential role in localization, the N-terminal third of the sequence has also been postulated to be important for sugar selectivity and channel formation (14, 29) and is highly conserved in enteric organisms. Residues 40 through 60 of LamB in *Escherichia coli*, *Salmonella typhimurium*, *Shigella* spp., and *Klebsiella pneumoniae* are identical except for one substitution (29). Residues 40 to 49 also constitute part of a conserved region identified by Nikaido and Wu (22) (Fig. 1) in outer membrane proteins such as LamB, OmpA, and OmpF. Hence, this region is potentially a more generally important structural feature in outer membrane proteins, although this conclusion has been questioned in studies of OmpA deletion mutants (13). However, no detailed genetic point mutation analysis of these sequences has been undertaken, and this study concentrates on amino acid replacements to test whether the nature of particular residues is significant in LamB biogenesis and function.

The detailed three-dimensional structure of LamB is not yet available, but a model of folding across the outer membrane has been derived, as illustrated in Fig. 1 (4). This model predicted that residues 40 to 53 constitute an amphipathic β-strand and residues 54 to 60 are part of a surface-exposed loop. A second aim of this mutagenesis study was to test the secondary structure predictions and indicate whether residues 54 to 60 are indeed more flexible in accepting substitutions than the predicted transmembrane segment involving residues 40 to 53.

MATERIALS AND METHODS

Strains, phages, and plasmids. *E. coli* K-12 strains were used in this study. Strain pop6510 [F$^-$ *thr leu metA lacY tonA supE recA56 srl*::Tn*10 lamB* (dex-5) (2)] has a chromosomal *lamB* null mutation and was used as the host strain for characterization of plasmid phenotype; BW2800 [*uraD139* Δ(*argF-lac*)*U169 rpsL relA* (*thi*) *ptsF25 flbB deoC1 rbsR malB* Δ(*malK-lamB*)*15 zja*::*Tn10*/F′(*lacI*^{q} *lacZ*::Tn*5 lacY*$^+$ *lacA*$^+$) (10)] was used in single-stranded DNA synthesis; CJ236 [*dut-1 ung-1 thi-1 relA1*/F′ CJ105(Cm^r) (16)] was used in the synthesis of uracilated single-stranded DNA for mutagenesis. Helper phage M13rv1 (30) was used in the syn-

* Corresponding author.

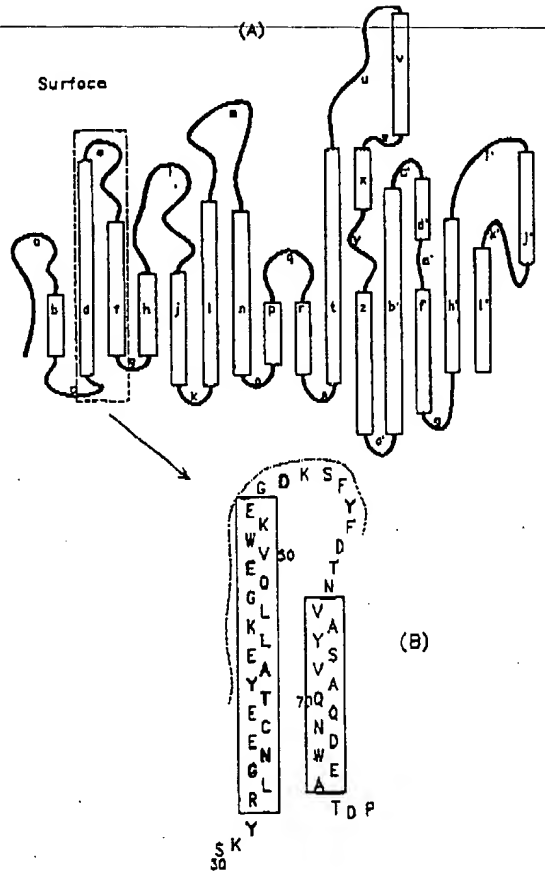

FIG. 1. (A) Predicted structure of maltoporin folding across the outer membrane (4), with alternating loosely structured regions (thick lines) and regions of ordered secondary structure (boxes) shown. (B) The region of study containing segments d and e is highlighted; residues 40 through 60 (dotted line) cover part of the second transmembrane region as well as residues 54 to 63, which were predicted to form a loop on the surface of the outer membrane. Identical residues in the corresponding region of OmpF are printed in boldface letters.

thesis of single-stranded DNA template for both mutagenesis and DNA sequencing. Bacteriophage λ_{vir} was used in lambda sensitivity assays by cross-streaking.

Plasmid pAM1850 was constructed from pAM1520 (10) by introducing a *Sac*I site covering the codon at residue 43 and a *Hin*dIII site at residue 56; neither of these DNA changes modified the amino acid coding sequence. Strain pop6510 harboring pAM1850 exhibited the same wild-type phenotype as with pAM1520 and was used as a positive control in all LamB function characterizations. pAM1854 encodes a stop codon (TAA) at residue 56, introduced by oligonucleotide-directed mutagenesis of pAM1850. pop6510 containing pAM1854 was used as the negative control in phenotypic characterizations.

Media and genetic techniques. LB and DYT media were used as described before (6). Minimal medium A (MMA) was prepared as described elsewhere (19). Eosin-methylene blue medium (EMB) was also used (19) and contained 0.4% maltooligosaccharide (Pfanstiehl Chemicals, Waukegan, Ill.) together with 0.04% eosin Y and 0.0065% methylene blue as pH indicators. For growth of plasmid strains, ampicillin was present at 50 μg/ml in MMA and LB and 100 μg/ml in DYT and EMB. All plasmid preparations were performed by the boiling method (7). Transformation of plasmids into bacteria and all other genetic techniques were performed as described before (18). Helper phage-aided single-stranded DNA synthesis was performed as described elsewhere (30).

Design and synthesis of degenerate oligonucleotides. Degenerate oligonucleotides 62 bases long, corresponding to the wild-type sequence 5′ CT TAT GCT GAG CTC AAA TTG GGT CAG GAA GTG TGG AAA GAG GGC GAT AAA AGC TTC TAT TTC 3′ (residues 40 to 60 of mature LamB as encoded in pAM1850), were chemically synthesized (Applied Biosystems). Each of the four nucleoside phosphoramidite substrates was contaminated with 2.33% of each of the other three nucleoside phosphoramidites. The proportion of oligonucleotides carrying substitutions was calculated as described before (21). With 7% contamination (i.e., 2.33% of each of the other three nucleoside phosphoramidites) in the synthesis of a 62-mer, we expected that 1.1% of the oligonucleotides obtained would have a wild-type sequence, 5.2% would have one base change, 12% would have two base changes, 17.9% would have three base changes, 20% would have four base changes, and 43.8% would have five or more base changes. The pool of synthesized oligonucleotides was purified by using an oligonucleotide purification cartridge (Applied Biosystems) to separate incomplete oligonucleotides.

Mutagenesis of residues 40 to 60. Plasmid pAM1850 was transformed into CJ236 for synthesis of the uracilated single-stranded DNA template. To obtain the Dex$^-$ template, mutagenesis of pAM1850 was performed with the degenerate oligonucleotides as described by Kunkel et al. (16), with *dut ung* selection, except that 30 ng of oligonucleotides and 1 μg of template were used in the annealing step. The uracilated template of pAM1854, which has a Lys-56→TAA (stop) mutation, was used in the second round of mutagenesis, again with the degenerate oligonucleotide pool. The ligation mixture was transformed into pop6510 (Dex$^-$) and spread onto 0.4% maltodextrin–EMB indicator plates. Clones which appeared dark red were picked and purified on nutrient agar plates.

Lambda and starch binding assays. Lambda sensitivity was assayed by cross-streaking isolated Dex$^+$ transformants against λ_{vir}.

The sugar-binding site of maltoporin was assayed in two ways. For screening of transformants, the chemotaxis soft agar plate (14) was used. Each of the selected Dex$^+$ isolates was tested for the ability of bacteria to bind to starch at a concentration of 2 mg/ml in 0.24% microbiological agar with 0.002% ribose present as an attractant. The size of the chemotaxis ring formed after incubation at 30°C overnight was recorded. Ring formation is indicative of a starch-binding defect, as starch prevents the swimming of bacteria with a wild-type level of functional sugar-binding site in the outer membrane (14). A more quantitative measure of the starch-binding ability of the isolates was made with washed suspensions of bacteria applied to starch-Sepharose columns as previously described (8); the proportion of bacteria retained in these columns is dependent on maltoporin-binding activity.

DNA sequence analysis. Double-stranded DNA sequencing was performed by standard Sequenase reactions as described by the manufacturer (US Biochemicals Corp., Cleveland, Ohio) except that 1 ng of primer was used for annealing, which was done by incubation at 37°C for 20 min.

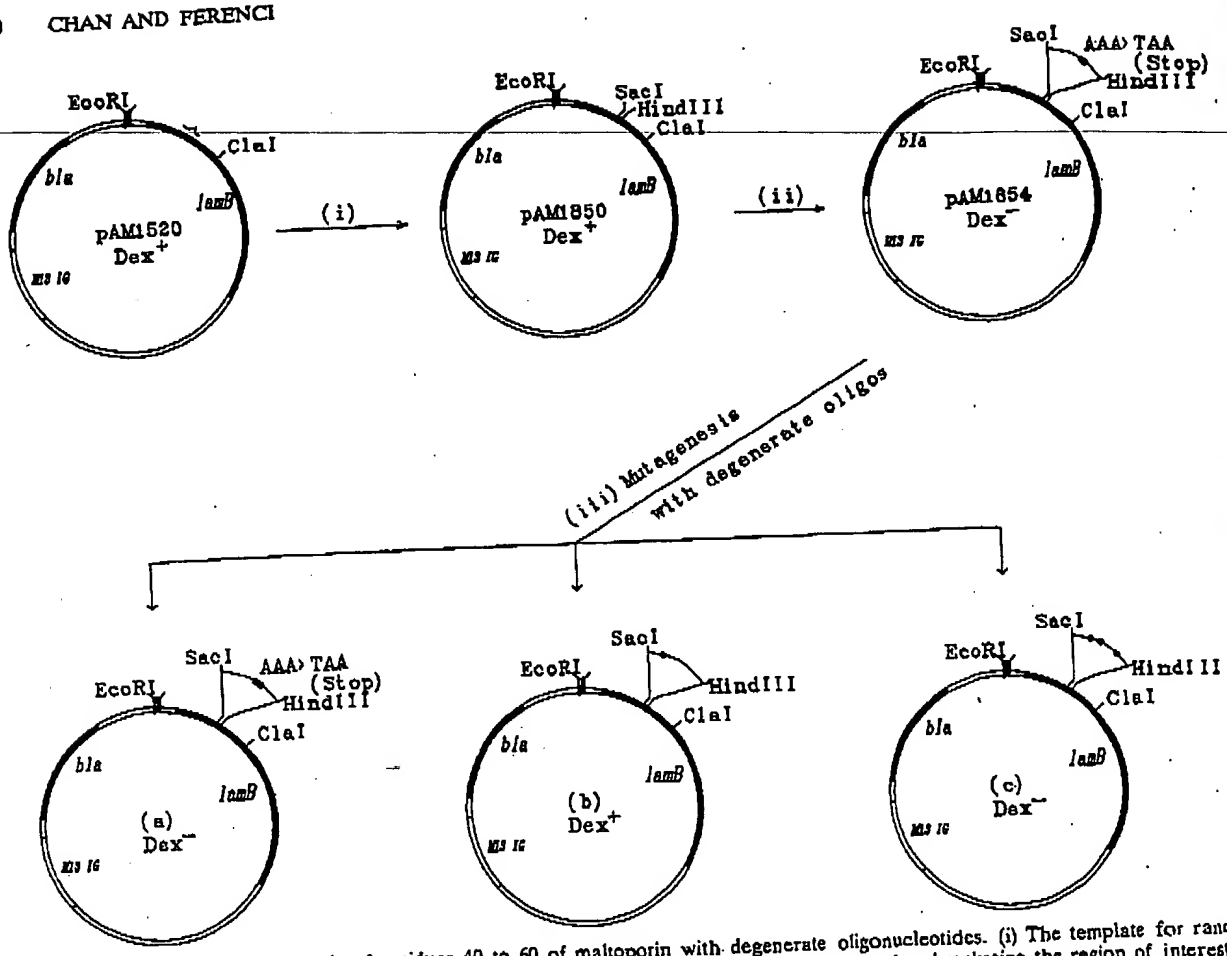

FIG. 2. Strategy for mutagenesis of residues 40 to 60 of maltoporin with degenerate oligonucleotides. (i) The template for random mutagenesis was constructed from plasmid pAM1520 (10) by inserting two unique restriction sites bracketing the region of interest by site-directed mutagenesis (16), forming pAM1850 (these restriction sites were not utilized in this study). The nucleotide changes in pAM1850 did not change the amino acid sequence of LamB. (ii) Degenerate oligonucleotides covering the region from residues 40 to 60 were used to introduce a stop codon within this region, as shown in pAM1854, resulting in loss of maltoporin function (Dex^-). (iii) The same pool of degenerate nucleotides were used to mutagenize pAM1854, with conventional *dut ung* selection. Three classes of transformants were expected to be obtained in this experiment: (a) transformants still carrying plasmids with the original stop codon in the region, resulting in a Dex^- phenotype; (b) transformants in which the stop codon region was replaced by an oligonucleotide sequence whose expression contributed to a pore-forming protein, with a resulting Dex^+ phenotype; and (c) transformants carrying plasmids with the region replaced by an oligonucleotide sequence whose expression results in a nonfunctional maltoporin, also resulting in a Dex^- phenotype. Only Dex^+ mutants were analyzed in this study; this selection ensures that all transformants investigated had their 40 to 60 region derived from a mutagenic oligonucleotide.

Sequencing reactions were carried out as soon as annealing was complete. The preparation of the double-stranded plasmid for sequencing was done as described before (7), but the following adjustments were made: 18 μl of double-stranded plasmids was denatured by adding 3.8 μl of 2 M NaOH and then incubated for 10 min at room temperature; 4.5 μl of 3 M sodium acetate (pH 4.8) was then added for neutralization. Single-stranded sequencing with phagemid DNA (30) was done when cross-banding problems occurred in the region.

Estimation of LamB levels in the outer membrane by gel electrophoresis. To quantify the amount of LamB in the outer membrane, outer membrane was extracted as described before (27) and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (17). Samples were boiled for 5 min before loading. Selected samples were also heated at 70°C for 5 min before loading to determine the stability of the mutant maltoporin trimers. For densitometric comparisons, Coomassie blue-stained gel tracks were scanned in an LKB 2202 laser densitometer.

RESULTS

A requirement in combinatorial mutagenesis with oligonucleotides is the recovery of large numbers of mutants altered at a target sequence and few isolates with wild-type or template sequences at the target. Also, the target should not be limited to regions flanked by restriction sites, as in cassette mutagenesis. Our general strategy for achieving these aims is shown in Fig. 2. The strategy involved initial isolation of a nonsense mutation at a codon within region for the residues 40 to 60 of LamB protein to give a template with a null phenotype. This was achieved by mutagenesis of a wild-type *lamB* (pAM1850) template with a pool of degenerate mutagenic oligonucleotides and screening for Dex^-

transformants on indicator plates. Thirteen of 203 transformants tested were unable to utilize maltodextrins (Dex⁻ phenotype). Three Dex⁻ clones were sequenced and were all found to contain a stop codon. One plasmid (pAM1854) altered at residue 56 (AAA→TAA) was used as the target template in subsequent experiments.

Mutagenesis of uracilated pAM1854 was performed with the same pool of degenerate oligonucleotides. Three possible outcomes could be predicted. Recovered transformants could have the same sequence as the template and be Dex⁻; these would be expected to constitute up to half of all transformants obtained with uracilated templates (16). Second, some transformants may have drastic changes in phenotype because of incorporation of oligonucleotides, including those from the pool with multiple substitutions, and would be Dex⁻ if unable to form a small number of channels in the outer membrane. It should be stressed that most substitutions affecting phage lambda or sugar binding are still Dex⁺, as are those causing even a fivefold drop in LamB levels in the outer membrane (14). The rest of the transformants would be Dex⁺, with at least some channel function but which could still be altered in phage or sugar binding. These transformant sequences would all be derived from the incorporation of mutagenic oligonucleotides, and only a small proportion of these were expected to have the wild-type sequence, given the level of misincorporation into the mutagenic oligonucleotide.

More than 1,500 isolates were obtained on maltodextrin-EMB-ampicillin plates, 290 of which, or approximately 20%, appeared dark red, indicating that maltodextrins were getting through the outer membrane. All 290 were found to be fully λ sensitive, suggesting that LamB trimer in the right conformation for phage binding was made in all these Dex⁺ isolates. Dex⁻ clones were also tested for λ sensitivity in an attempt to screen for those with a correct surface conformation but unable to utilize maltodextrins because of a drastic change in channel conformation. However, all 100 Dex⁻ clones cross-streaked were λ resistant, indicating that no functional trimer was present in the outer membrane of these Dex⁻ isolates. These Dex⁻ transformants were presumed to include a high proportion carrying the template stop codon at position 56, as well as some with mutations, including multiple mutations, rendering them unable to form functional protein. These isolates were not investigated further.

The Dex⁺ clones were also tested for the functional state of their sugar-binding site; previous studies showed that sugar affinity defects can be present in mutants that form a channel permitting growth on dextrins (14). Starch binding by the sugar-binding site was assayed by the soft agar-chemotaxis plate method, in which bacteria with functional binding sites are rendered immobile by interaction with starch (14). Of the 180 Dex⁺ clones tested, 23 were Bin⁻, 3 of which contained a stop codon (TAG) at residue 41, 48, and 59, respectively; these contained some protein because of the *supE* suppressor of the host strain pop6510, leading to a Dex⁺ phenotype with Gln at these sites. The remaining 20 Bin⁻ isolates were further characterized as shown below. The other 157 isolates were phenotypically wild type by the criteria of phage receptor function, sugar binding, and channel formation for dextrins.

The substitutions in 87 of the 157 Dex⁺ Bin⁺ isolates were determined by DNA sequencing. The proportion of the Bin⁺ isolates having a wild-type DNA sequence in the region was found to be 14.9%. Also, 40.2% of these isolates contained one base change compared with the wild type (i.e., *lamB* sequence as in pAM1850), 29.9% contained two base changes, 9.2% contained three, 3.4% contained four, and 2.3% contained five base changes. No isolates containing more than five base changes were obtained. Six of the 35 Bin⁺ isolates which contained one base change were actually wild type at the amino acid sequence level, and hence only 69 of the 87 Dex⁺ Bin⁺ isolates had amino acid replacements. When translated into amino acid changes, 53% of the Bin⁺ isolates contained one amino acid change, 32% contained two amino acid changes, 13% contained three amino acid changes, and 2% contained four or more amino acid changes. Also, we found little bias in the incorporation of nucleotides: there were 35 base changes to A, 48 base changes to T, 54 base changes to C, and 47 base changes to G.

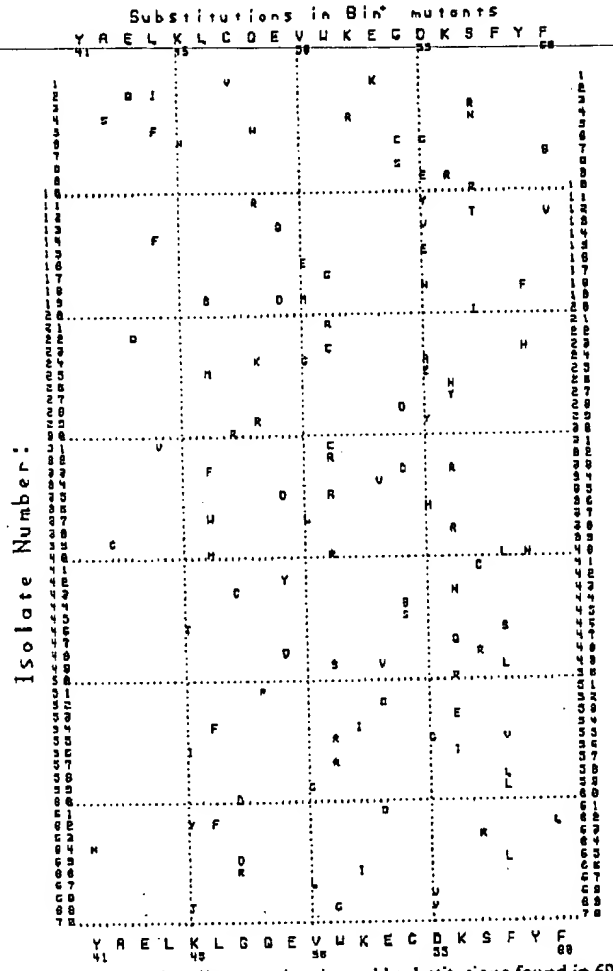

FIG. 3. Functionally neutral amino acid substitutions found in 69 isolates exhibiting a wild-type Dex⁺ Bin⁺ λˢ phenotype. Amino acids are represented by the single-letter amino acid code. The top and bottom sequences are wild-type residues 41 through 60 of maltoporin.

A summary of the functionally neutral amino acid substitutions (excluding silent substitutions within a codon) found in each of the 69 Dex⁺ Bin⁺ isolates is given in Fig. 3. Phenotypically neutral substitutions were found at every residue except residue 40, at the extreme 5′ end of the mutagenized region. Several of the substitutions occurred

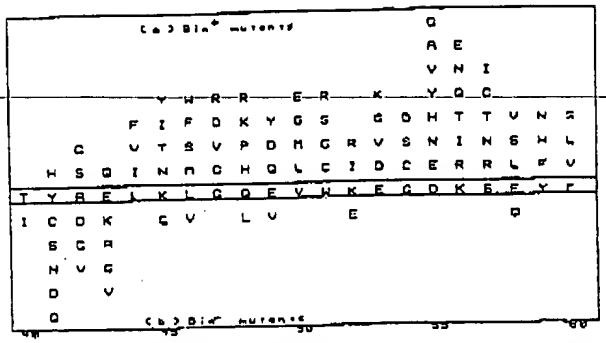

FIG. 4. Summary of functionally neutral and nonneutral amino acid substitutions recovered at each residue in the region containing residues 40 to 60 of maltoporin. The wild-type sequence is shown in the boxed middle line. (a) Above this line are amino acid substitutions that have undetectable effects on LamB functions. (b) Shown below the wild-type residues are substitutions recovered exclusively in the Dex$^+$ Bin$^-$ isolates. Each type of substitution is shown once even if it occurred in several of the isolates shown in Fig. 3 or in different combinations with other substitutions.

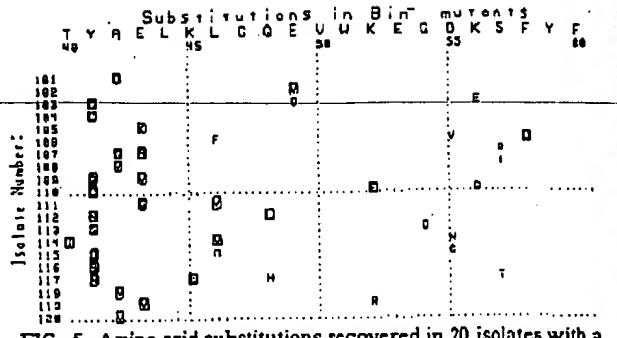

FIG. 5. Amino acid substitutions recovered in 20 isolates with a starch-binding defect (phenotypically Dex$^+$ Bin$^-$ λ^s). The wild-type sequence of residues 40 to 60 of maltoporin is shown on the top line, followed by the amino acid replacements of each Dex$^+$ Bin$^-$ isolate. Residue changes unique to Bin$^-$ mutants are boxed (i.e., if not found among the neutral substitutions in Fig. 3).

more than once and, in double mutants, were often found in random combinations with other substitutions. For example, the Trp-51→Arg substitution was found in six isolates, three times by itself (isolates 21, 32, and 57) and once each with one, two, or three other replacements (isolates 35, 40, and 55, respectively; Fig. 3). It is therefore unlikely that these multiple substitutions are required to interact to give a Dex$^+$ phenotype, and we assume that each of the substitutions is functionally neutral. This conclusion is similar to that previously obtained in a mutagenesis study of a well-characterized protein (3). As a simplified summary, all possible neutral replacements for each residue, even when found in combination with other replacements, are listed in Fig. 4a.

After sequencing 87 Dex$^+$ Bin$^+$ isolates, it became apparent that the same substitutions were reappearing in the sequenced clones. The pattern of substitutions in Fig. 3 may have been biased by oligonucleotide hybridization preference, repair at particular codons, or other factors unrelated to the significance of particular amino acids to LamB function. One way of testing whether the lack of neutral substitutions at residues 41 to 43 was a mutagenesis artifact was to sequence defective mutants as well. If the lack of substitutions in Dex$^+$ Bin$^+$ mutants was a mutagenesis artifact, then there should also be few defective mutations in the 40 to 43 segment.

Twenty Dex$^+$ Bin$^-$ clones were sequenced, and the substitutions shown in Fig. 5, including the combinations of multiple substitutions, were found. Most strikingly, the replacements uniquely present among Bin$^-$ isolates were mostly at residues 40 to 43, although some of these were in random combination with substitutions already found in neutral Dex$^+$ Bin$^+$ isolates. As summarized in Fig. 4b, in which only the uniquely Bin$^-$ substitutions are noted, there is a satisfying inverse correlation with the pattern of the Bin$^+$ clones. Sites with few functional Dex$^+$ Bin$^+$ substitutions were most frequently substituted among the Bin$^-$ isolates, indicating that the lack of neutral substitutions at residues 41 to 43 was not a mutagenesis artifact.

It should be noted that the Dex$^+$ Bin$^-$ phenotype in the chemotaxis assay was previously found to be due either to a reduced amount of protein in the outer membrane or to normal amounts of protein but reduced sugar affinity at the sugar-binding site (14). To distinguish between these possibilities, the levels of LamB in the outer membrane of selected Bin$^+$ and Bin$^-$ isolates were estimated. Outer membrane was extracted from 18 nonidentical Bin$^-$ isolates, and the levels of LamB in the outer membrane were determined by gel electrophoresis, as shown in Fig. 6. The results indicated that all of the Bin$^-$ isolates have some reduction in the level of LamB protein in the outer membrane, as summarized in Table 1, in which densitometrically scanned LamB levels were normalized to OmpF/C levels to compensate for possibly misleading loading differences in Fig. 6. Also in contrast to wild-type protein trimers, which are stable and hardly dissociate at 70°C, four (isolates 105, 107, 113, and 115) of five isolates tested had LamB that dissociated more readily at the lower temperature (Table 1). In particular, LamB from isolate 105 (Glu-43→Lys) was completely dissociated, suggesting that trimer stability was affected by Bin$^-$ mutations. It also appeared that these mutant LamB trimers are more prone to proteolysis (results not shown). Hence, residues 40 to 43 are important for maltoporin structure.

To further analyze the effects of mutations on the sugar-binding site in Bin$^-$ isolates, starch binding was also assayed in starch-Sepharose columns. Ninety percent of the bacteria with a wild-type LamB-expressing plasmid (pop6510 harboring pAM1850) were retained in these columns, whereas only 10% of the bacteria without LamB were retained. For the mutant isolates, the retention rate, expressed as the percentage of bacteria retained in the column, was as summarized in Table 1. In confirmation of the chemotaxis experiments, all

FIG. 6. LamB protein in the outer membrane analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (17). Track a shows the pattern of outer membrane proteins of pop6510 containing pAM1850; tracks b through s show Bin$^-$ isolates 101 through 118, respectively. Purified protein standards were in track t. MBP, maltose-binding protein.

TABLE 1. Summary of protein levels and binding activity of Bin⁻ isolates

Isolate no.	Substitution[a]	Normalized LamB level in outer membrane[b] (% of wild-type level)	Dextrin binding[c] (% retention by starch-Sepharose)	% Dissociation of LamB[d]
Wild type		100	90	7
101	A42D	27	40	ND[e]
102	E49V	16	35	ND
103	Y41C	0	1	ND
104	Y41C	16	16	ND
105	E43K	21	8	100
106	F58D	3	26	ND
107	A42G, E43A	40	47	62
108	A42V	80	48	ND
109	Y41Q, E43V	29	2	ND
110	Y41C, K52E	6	31	ND
111	E43G, L46V	22	22	ND
112	Y41N, Q48L	26	1	ND
113	Y41S	30	8	53
114	T40I, L46V	82	45	ND
115	Y41D	22	20	37
116	Y41S	32	32	ND
117	Y41N, K45E	27	21	ND
118	A42V	56	26	9
119	E43V	ND	25	ND
120	A42V	ND	28	ND
No LamB		0	10	ND

[a] Substitutions unique to those found in the Bin⁻ isolates are tabulated; the complete set of substitutions in each isolate are shown in Fig. 5.

[b] The gel in Fig. 6 was scanned densitometrically, and the ratio of peak sizes of LamB relative to that of the combined OmpF/C bands was estimated for each isolate. The value given is the LamB/OmpFC percentage for isolates relative to that for the wild type.

[c] Starch binding of bacteria containing LamB-expressing plasmids was estimated as described previously (14).

[d] Outer membrane extracts from the isolates were electrophoresed after heating at 70 and 100°C. The stained LamB bands were scanned densitometrically, and the proportion of the LamB monomer at 70°C versus that at 100°C was tabulated.

[e] ND, not determined.

Bin⁻ isolates had reduced starch-binding ability. This was also true for those mutants (isolates 108, 114, and 118) that had only slightly reduced levels of LamB in the outer membrane (Table 1). Hence, some substitutions may influence the conformation of the sugar-binding site as well as change protein levels.

DISCUSSION

The strategy used in this study has a number of advantages over current combinatorial cassette mutagenesis methods. By forcing the mutagenic oligonucleotides to replace a nonsense codon in the template, it is possible to ensure that all functional or partly functional sequences result from the oligonucleotides, regardless of the rate of mutagenesis. This is simpler than the cassette approach (3), which requires unique flanking restriction sites for the efficient cloning-in of large numbers of mutagenic sequences. The use of a uracilated template also reduced the frequency of template sequences among our transformants, and our 20% rate of recovery of Dex⁺ transformants is reasonably satisfactory. Among the Dex⁺ clones sequenced, there was a preferential incorporation of oligonucleotides carrying three or fewer mismatches. This is itself an advantage in that the Dex⁺ isolates were limited in the number of substitutions they carried. However, a limitation of this method, common to all mutagenesis methods for extended sequences, was that total randomization of each codon in the region was not feasible. Hence, the spectrum of substitutions observed was generally biased in favor of codons differing by a single base from the wild-type codon. Nevertheless, we recovered seven mutants carrying two base changes in the same codon, and thus by sequencing a large number of mutants, a wide spectrum of functionally acceptable and defect-causing mutations, encoding a range of phenotypes, could be obtained.

Two approaches were used in the analysis of mutants. First, functionally neutral substitutions at each residue in the region were identified to permit identification of unimportant residues. The second approach involved identifying sites of mutations present in isolates with a LamB defect. The results obtained with these two approaches showed a good negative correlation; sites with few permitted neutral substitutions were more frequently sites of deleterious substitutions and vice versa. Hence, we have confidence in the conclusions on the relative informational importance of residues in the 40 to 60 region of maltoporin. We conclude that residues 40 to 60 of LamB can be divided into four stretches of distinct functional properties.

Residues 40 to 43. Residues 40 to 43 show homology to several outer membrane proteins (22). It is significant that most mutants with a starch-binding defect and reduced levels of protein in the outer membrane had substitutions at residues 41 to 43. No functional substitutions were recovered at residue 40, but this is probably because only the second and third bases of the codon were present at the extreme 5' end of the mutagenic oligonucleotide. However, the recovery of a Bin⁻ isolate carrying a Thr-40→Ile mutation suggests that this residue may also be significant in maltoporin function. The results of starch-binding assays and protein level assays (Fig. 6 and Table 1) with the Bin⁻ isolates indicated that mutations at residues 40 to 43 led to a severe defect in both the level and stability of maltoporin in the outer membrane, suggesting that these residues, and residues in the corresponding region in other outer membrane proteins, play an important role in the localization and assembly processes. A few amino acid substitutions influenced starch binding without greatly reducing protein levels in the outer membrane. Given the structural changes elicited by other substitutions in the 40 to 43 region, it is likely that the sugar-binding defect is an indirect conformational effect on the binding site rather than a change directly at the binding site itself.

Since the completion of this study, the position of residues in OmpF corresponding to residues 41 to 43 of LamB has been determined by X-ray crystallography (5). These residues were in a transmembrane region deep within the monomer-monomer interaction site of OmpF. The phenotype of LamB mutants within this region is highly consistent with such a position in LamB as well.

Residues 55 to 57. In contrast to residues 40 to 43, residues 55 to 57 accept a diverse range of substitutions (Fig. 4a), suggesting that these residues are unlikely to be of functional importance. The flexibility demonstrated at these residues is in good agreement with the proposed secondary-structure model (Fig. 1) (4), which suggested that these residues constitute a membrane-external loop. It has previously been demonstrated that a Ser-57→Cys mutant had phenotypes no different from those of the wild type (12), which is also consistent with the tolerance of these residues in accepting neutral substitutions. However, labeling studies on the mutant carrying the Ser-57→Cys mutation showed that the thiol

was not as accessible as expected for a freely external loop (11). In OmpF, there is no membrane-external loop but a tight beta turn at the corresponding position. There may well be a difference from LamB at this site.

Residues 58 to 60. The region of the highest flexibility (residues 55 to 57) does not extend to residues 58 to 60, which were also postulated to be in the loop. A mutant with a substitution of Phe-58→Asp (together with two other neutral substitutions at residues 46 and 55) had a starch-binding defect. Although at this stage we have no data on the effect of the Phe-58→Asp mutation in isolation, the relatively restricted nature of residues 58 to 60 suggests that these residues are not part of the flexible loop. The mutational data are more consistent with these residues already being in the next ordered segment (f, Fig. 1).

Residues 44 to 54. Residues 44 to 54, with the exception of residues 44, 49, 52, and 54, all accept a limited range of functional substitutions (Fig. 3). At this stage, it is difficult to conclude whether this stretch of amino acid residues is part of a transmembrane region, as in Fig. 1, or a region such as one of the short α-helices found in OmpF or *Rhodobacter* porin (5, 27). Alternatively, residues 44 to 54 together with residues 55 to 57 could constitute a larger external loop. A larger loop would require a shorter preceding transmembrane segment, and given that most transmembrane β-structured segments in OmpF or *Rhodobacter* porin are less than 17 residues in length (5, 27), the originally proposed 21 residues in segment d (Fig. 1) may have been overestimated.

Nevertheless, the nature of some residues between 44 and 54 appears to be critical. At residue 44, only large and hydrophobic amino acids were recovered as neutral substitutions (Fig. 3). Also, Glu-49→Val resulted in a starch-binding defect (Fig. 5). Given that this residue accepts only relatively conservative substitutions and that the same residue is found in OmpF (Fig. 1), it appears that Glu-49 plays a role in the stability and functioning of LamB in the outer membrane. Similarly, residues 52 and 54 also appear to have a restricted range of functional substitutions. In particular, Lys-52 accepts only Arg, which is of the same charge, and Ile, which is identical to the corresponding residue in OmpF. Also, a Lys-52→Glu substitution, along with two other substitutions at residues 41 and 56, resulted in a Bin$^-$ phenotype (isolate 110, Fig. 5), although the effect of the Lys-52→Glu substitution in isolation is not known at this stage.

Since both residues 49 and 52 are charged residues, further investigations are necessary to elucidate whether their potential roles in LamB structure and function are charge related. It is important to note that, as shown in Fig. 1, segment d is highly polar for a potential transmembrane segment and also shows high amphipathicity as a β-structure, as is the corresponding region of OmpF (5). Substitutions changing the nature of the charge were not tolerated at all except in the loop at residues 53 to 56, as were other nonconservative substitutions. Charge changes at residues 45, 49, and 52 were associated with a protein structural defect. These results are consistent with the likely importance of the amphipathic nature of this segment in LamB structure, stability, and outer membrane assembly.

In conclusion, the results presented in this study have proved useful in the identification of important residues within a region for which little structural and functional information was available. The spectrum of neutral and defective substitutions has provided an internally consistent pattern of the particular significance of residues 40 to 43 in the structure and function of LamB and possibly in the corresponding regions of other outer membrane proteins. It is also worth emphasizing that the study of functionally acceptable substitutions provides a means of testing proposed models of secondary structure.

ACKNOWLEDGMENT

This work was supported by grants from the Australian Research Committee (to T.F.).

REFERENCES

1. **Benz, R., G. Francis, T. Nakae, and T. Ferenci.** 1992. Investigation of the selectivity of maltoporin channels using mutant LamB proteins: mutations changing the maltodextrin binding site. Biochim. Biophys. Acta **1104**:299–307.
2. **Bouges-Bocquet, B., H. Villaroya, and M. Hofnung.** 1984. Linker mutagenesis in the gene of an outer membrane protein of *Escherichia coli*, LamB. J. Cell. Biochem. **24**:217–228.
3. **Bowie, J. U., and R. T. Sauer.** 1989. Identifying determinants of folding and activity for a protein of unknown structure. Proc. Natl. Acad. Sci. USA **86**:2152–2156.
4. **Charbit, A., K. Gehring, H. Nikiado, T. Ferenci, and M. Hofnung.** 1988. Maltose transport and starch binding in phage resistant point mutants of maltoporin: functional and topological implications. J. Mol. Biol. **201**:487–493.
5. **Cowan, S., T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and J. P. Rosenbusch.** 1992. Crystal structures explain functional properties of two *E. coli* porins. Nature (London) **358**:727–733.
6. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Del Sal, G., G. Manfioletti, and C. Schneider.** 1988. A one-tube plasmid DNA minipreparation suitable for sequencing. Nucleic Acids Res. **16**:9878.
8. **Ferenci, T., and K.-S. Lee.** 1982. Directed evolution of the lambda receptor of *Escherichia coli* through affinity-chromatographic selection. J. Mol. Biol. **160**:431–444.
9. **Ferenci, T., and T. J. Silhavy.** 1987. Sequence information required for protein translocation from the cytoplasm. J. Bacteriol. **169**:5339–5342.
10. **Ferenci, T., and S. Stretton.** 1989. Cysteine-22 and cysteine-38 are not essential for the functions of maltoporin (LamB protein). FEMS Microbiol. Lett. **61**:335–340.
11. **Francis, G., L. Brennan, and T. Ferenci.** 1991. Affinity-chromatographic purification of sixteen cysteine-substituted maltoporin variants: thiol reactivity and cross-linking in an outer membrane protein of *Escherichia coli*. Biochim. Biophys. Acta **1067**:89–96.
12. **Francis, G., L. Brennan, S. Stretton, and T. Ferenci.** 1991. Genetic mapping of starch and lambda receptor sites in maltoporin: identification of substitutions causing direct and long-range effects on binding sites by cysteine mutagenesis. Mol. Microbiol. **5**:2293–2301.
13. **Freundl, R., H. Schwarz, M. Klose, N. Rao Move, and U. Henning.** 1985. The nature of information required for export and sorting, present within the outer membrane protein OmpA of *Escherichia coli*. EMBO J. **4**:3593–3598.
14. **Heine, H.-G., G. Francis, K. S. Lee, and T. Ferenci.** 1988. Genetic analysis of sequences in maltoporin that contribute to binding domains and pore structure. J. Bacteriol. **170**:1730–1738.
15. **Hinkle, P. C., P. V. Hinkle, and H. R. Kaback.** 1991. Information content of amino acid residues in putative helix VIII of the lac permease from *Escherichia coli*. Biochemistry **29**:10989–10994.
16. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
17. **Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen.** 1975. Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K12 into four bands. FEBS Lett. **58**:254–258.

18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Misra, R., A. Peterson, T. Ferenci, and T. J. Silhavy.** 1991. A genetic approach for analysing the pathway of LamB assembly into the outer membrane of *Escherichia coli*. J. Biol. Chem. 266:13592–13597.
21. **Ner, S. S., and M. Smith.** 1989. Targeted random mutagenesis using "doped" synthetic oligonucleotides. Mol. Biol. Rep. (Bio-Rad Laboratories) 7:1–2.
22. **Nikaido, H., and H. C. P. Wu.** 1984. Amino acid homology among the major outer membrane proteins of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81:1048–1052.
23. **Oliphant, A. R., and K. Struhl.** 1989. An efficient method for generating proteins with altered enzymatic properties: application to β-lactamase. Proc. Natl. Acad. Sci. USA 86:9094–9098.
24. **Rasmussen, B. A., and T. J. Silhavy.** 1987. The first 28 amino acids of mature LamB are required for rapid and efficient export from the cytoplasm. Genes Dev. 1:185–196.
25. **Reidhaar-Olson, J. F., J. U. Bowie, R. M. Breyer, J. C. Hu, K. L. Knight, W. A. Lim, M. C. Mossing, D. A. Parsell, K. R. Shoemaker, and R. T. Sauer.** 1991. Random mutagenesis of protein sequences using oligonucleotide cassettes. Methods Enzymol. 208:564–568.
26. **Reidhaar-Olson, J. F., and R. T. Sauer.** 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241:53–57.
27. **Schiltz, E., A. Kreusch, U. Nestel, and G. E. Schulz.** 1991. Primary structure of porin from *Rhodobacter capsulatus*. Eur. J. Biochem. 199:587–594.
28. **Schnaitman, C. A.** 1971. Solubilization of the cytoplasmic membrane of *Escherichia coli* by Triton X-100. J. Bacteriol. 108:545–552.
29. **Werts, C., A. Charbit, S. Bachellier, and M. Hofnung.** 1992. DNA sequence analysis of the *lamB* gene from *K. pneumoniae*. Implications for the topology and the pore functions in maltoporins. Mol. Gen. Genet. 233:372–378.
30. **Zagursky, R. J.** 1985. pBR322 plasmids that yield single-stranded DNA. BRL Focus 7:4–6.

Gene, 106 (1991) 7–12

GENE 06073

Combinatorial mutagenesis of three major groove-contacting residues of *Eco*RI: single and double amino acid replacements retaining methyltransferase-sensitive activities

(DNA-protein interactions; enzyme specificity; polymerase chain reaction; protein engineering)

Joel Osuna, Humberto Flores and Xavier Soberón

Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos 62271 (Mexico)

Received by F. Bolivar: 6 May 1991
Revised/Accepted: 20 May/23 July 1991
Received at publishers: 29 July 1991

SUMMARY

A library of mutant *ecoRIR* genes encoding *Eco*RI restriction endonuclease was generated using trinucleotide blocks and a combination of recombinant DNA procedures, including primer extension and the polymerase chain reaction. Codons corresponding to three amino acids (E^{144}, R^{145} and R^{200}), previously implicated in the specific recognition of the DNA substrate, were combinatorially mutated so as to generate a library that potentially contains all 20^3 possible single, double and triple aa replacements, in a balanced distribution. Inspection of the phenotypes of *Escherichia coli* colonies bearing the mutant genes showed that several of them retained activities that were deleterious to the cells but were still protected by the *Eco*RI methyltransferase. These included new enzyme variants, including non-conservative single (Thr or Val for Glu^{144}) and double (Val for Glu^{144} and Thr for Arg^{145}) replacements.

INTRODUCTION

In the study of the structure–function relationship of proteins, as well as for the generation of proteins with novel properties, methods for the site-directed mutagenesis of the corresponding genes have become an indispensable tool. There are several stages, conceptual and methodological, to be taken into account when designing a rational mutagenesis strategy: first, even when there is a known three-dimensional structure for the protein, it is extremely difficult to predict which changes are necessary to achieve the desired property of the protein. This type of prediction is still harder in the event that more than one residue change is needed. Second, frequently the residues one wishes to alter are far away from each other in the primary sequence (although they may be close together in the active conformation of the protein), and the combinatorial alteration of such residues is highly desirable (Wells, 1990). Third, as the number of target residues increases, the use of codons as mutagenic units becomes important in order to obtain a large number of useful alleles within a manageable mutant population (Sirotkin, 1986). We propose here a method for combinatorial, saturation mutagenesis, with codons as mutagenic units.

In order to test a mutagenesis method incorporating the considerations mentioned above, we chose the restriction endonuclease *Eco*RI as a model protein. The first report on the structure of the protein-DNA complex, to 3Å resolution

Correspondence to: Dr. X. Soberón, CIIGB/UNAM, Apdo. Postal 510-3, Col. Miraval 62271, Cuernavaca, Morelos (Mexico)
Tel. (52-73) 172-399; Fax (52-73) 172-388.

Abbreviations: aa, amino acid(s); bp, base pair(s); *ecoRIR*, gene encoding *Eco*RI; *ecoRIM*, gene encoding M·*Eco*RI; ENase, restriction endonuclease; EtdBr, ethidium bromide; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, 1000 bp; Km, kanamycin; LB, Luria–Bertani (medium); M·*Eco*RI, *Eco*RI MTase; MTase, methyltransferase; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; ^{R}, resistance; ss, single-strand(ed); wt, wild type.

(McClarin et al., 1986), led to the proposition that three aa (E^{144}, R^{145} and R^{200}) directly contact the purine bases at the major groove of the substrate DNA molecule. A recent revision of the structure (Kim et al., 1990) reveals additional regions of the protein that contact the pyrimidine bases, also at the major groove, in a new model which involves differences in chain connectivity. The lethal activity of the enzyme in vivo, together with the existence of a corresponding MTase, which inhibits DNA cleavage by the ENase, has permitted the design of sufficiently sensitive in vivo assays for the detection of mutant ENases with either normal or altered activity and/or specificity for the canonical site (Heitman and Model, 1990; Oelgeschläger et al., 1990).

In this paper, we report the generation of a library of DNA fragments, potentially containing all of the 8000 possible aa combinations at three positions that face the DNA major groove (E^{144}, R^{145} and R^{200}). We used trinucleotide blocks in the oligo synthesis and obtained a fragment library which resulted in the expected distribution of replacements (single, double and triple), with no obvious bias and no stop codons. Using this library and a colony phenotype assay, we have isolated new partially active mutants, including a double aa replacement, that further strengthen the notion of the importance of DNA conformation (Lesser et al., 1990), and/or additional contacts at the bases (Heitman and Model, 1990), on the recognition process.

RESULTS AND DISCUSSION

(a) Combinatorial mutagenesis of aa 144, 145 and 200

In order to have access to all 20^3 possible combinations of aa at these three positions, we designed an experiment in which we would introduce, via synthetic DNA, a fragment consisting of 50% wt codon, 50% of an equimolar mixture of codons for the 20 aa, at each of the three positions (see Fig. 1 legend). In the ideal case (e.g., with equal coupling efficiencies of the trinucleotides and no other artifacts during chemical and enzymatic manipulations), a binomial distribution would predict the composition of the fragment to be 12.5% wt in all three positions, 37.5% single, 37.5% double and 12.5% triple replacements. The number of clones sequenced so far (as shown in Table I) is still too low to tell how close the actual distribution of changes is, compared to the ideal. Nonetheless, our results do show that we have a collection with a variety of replacements, in all three positions, which has already provided mutants with interesting phenotypes (see section b).

This mutagenesis scheme goes beyond previous reports on combinatorial mutagenesis (Reidhaar-Olson and Sauer, 1988; Dunn et al., 1988; Sartorius et al., 1989). First, we have targeted residues that are far away from each other in the sequence. Second, through the use of trinucleotides we have aimed at a non-biased collection of replacements,

TABLE I

Variant enzyme phenotypes

*Eco*RI mutants[a]	aa[b]			Plating phenotype[c] without M · *Eco*RI[d]	
	E^{144}	R^{145}	R^{200}	−IPTG	+IPTG
wt	—	—	—	+	−
E144C	C	—	—	+ +	−
E144T	T	—	—	+ + + + +	+
E144V	V	—	—	+ + + + +	+
E144G	G	—	—	+ + + + +	+ + + + +
E144F	F	—	—	+ + + + +	+ + + + +
E144W	W	—	—	+ + + + +	+ + + + +
R145K	—	K	—	+ + + + +	+ + +
R200C	—	—	C	+ + +	+
R200K	—	—	K	+ + + + +	+ + + + +
R200N	—	—	N	+ + + + +	+ + + + +
R200H	—	—	H	+ + + + +	+ + + + +
R200Q[e]	—	—	Q	ND	ND
ER-VT[f]	V	T	—	+ + + + +	+ +
ER-KY	K	Y	—	+ + + + +	+ + + + +
ER-PF	P	F	—	+ + + + +	+ + + + +
ER-YW[e] (ASN 149)	Y	W	—	ND	ND
ER-DE[e]	D	—	E	ND	ND
ERR-VTG	V	T	G	+ + + + +	+ + + + +
ERR-IKL (LEU 203)	I	K	—	+ + + + +	+ + + + +
ERR-DIH[e]	D	I	H	ND	ND
ERR-RHT	R	H	T	+ + + + +	+ + + + +

[a] Designations for variant ENase indicate the original aa at the left and the aa present in the mutant, at the right, separated by the aa number where the replacement occurred. Numbers for multiple replacements are omitted for simplicity, they can be derived from columns 2–4. Numbers in parentheses denote mutations present outside the intended mutagenesis window.

[b] Sequencing was performed on the entirety of *Pst*I-*Hind*III fragments, derived from mutant colonies, cloned on M13mp19. The same fragments were also recloned in *Pst*I-*Hind*III-digested pKGS to verify the observed phenotypes (shown in Fig. 2).

[c] JM101 colony phenotypes were assessed visually. Colonies were grown on Km plates at a final concentration of 50 μg/ml. Where indicated, plates were supplemented with IPTG to a final concentration of 1 mM. + + + + +, normal appearing colony; + + + +, lower density colonies but similar size; + + + and + +, reduced size and visibly translucent colonies; +, flat, small and translucent colonies (similar appearance as the parent enzyme). −, no growth. ND, not determined.

[d] When the mutant genes were expressed in a MTase context, all colonies had a normal phenotype, including colonies bearing the parent *ecoRIR* gene. Under induction conditions only the parent gene is lethal (see Fig. 2).

[e] Data obtained from direct cloning of the *Bgl*II-*Pst*I fragment in M13 vectors.

[f] In this case, the *ecoRIM* gene was complemented both in *trans*, in a compatible plasmid, and in *cis*, with the *ecoRIM* gene cloned in the same plasmid as the *ecoRIR* gene. In both cases, there was a protective effect of the MTase on the lethality of the mutant ENase under induction conditions. The rescue was more evident in the cells containing the *ecoRIM* gene present in *trans* (see Fig. 2).

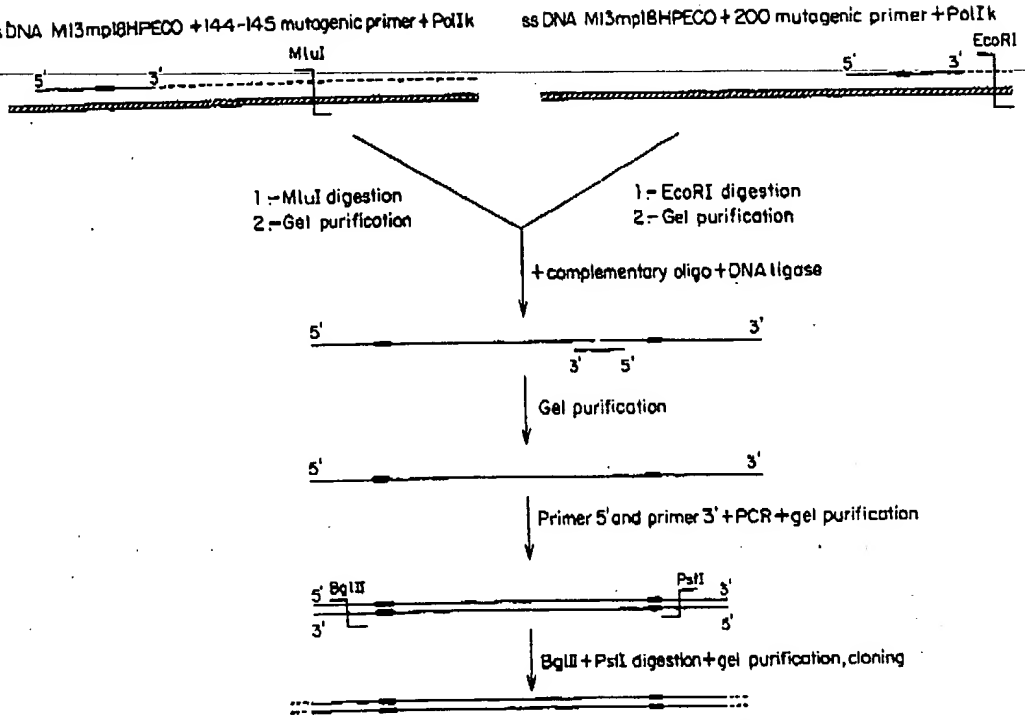

Fig. 1. Combinatorial mutagenesis strategy. The short, solid lines indicate synthetic oligos used as mutagenic primers or as ligation adapters (thickenings denote mutagenic regions). Hatched double lines denote the M13mp18 ss DNA template. Dashed lines depict the extension products. **Methods:** All cloning experiments were carried out with M13mp18, M13mp19 (Yanisch-Perron et al., 1985) or pKGS (Kuhn et al., 1986) as vector DNA and JM101 (Messing, 1979) as recipient strain. Plasmid pKGS, which carries a full length *ecoRIR* gene under the control of the inducible promoter lacUV5, and pSC101meth, a Tc^R derivative of pSC101, containing the *ecoRIM* gene (Betlach et al., 1976), were a generous gift from P.J. Greene (UCSF). Oligos were synthesized on a DNA synthesizer (Systec 1480A) using the phosphoroamidite method (Beaucage and Caruthers, 1981) and controlled pore glass as support (Adams et al., 1983). In some steps, those involving the introduction of a cocktail of trinucleotides, a mixture was prepared containing equimolar amounts of 20 different, fully protected, phosphotriester trinucleotides, synthesized as described (Broka et al., 1980). Each trinucleotide corresponded to a codon for one aa, as follows: Ala, GCT; Arg, CGT; Asn, AAC; Asp, GAC; Cys, TGT; Glu, GAA; Gln, CAG; Gly, GGT; His, CAC; Ile, ATC; Leu, CTG; Lys, AAA; Met, ATG; Phe, TTT; Pro, CCG; Ser, TCT; Thr, ACT; Trp, TGG; Tyr, TAC and Val, GTA. At the positions corresponding to codons selected for replacements, the growing oligo was reacted, using the procedure of Ito et al. (1982), through a manual injection valve, with a solution made with a total of 20 μmol trinucleotides from the cocktail described above and 20 μmol of the trinucleotide corresponding to the aa present in the wt *Eco*RI sequence. Purification of oligos was performed by gel electrophoresis in 20% polyacrylamide-8 M urea gels. Large pieces were sliced from the gel to allow for heterogeneity of migration of the mixed oligos. For the creation of convenient restriction sites, oligo-directed mutagenesis was performed essentially as described (Su and El-Gewely, 1988). M13mp19HPECO, an M13mp19 clone that carries the *Hin*dIII-*Pst*I insert, which corresponds to the fragment coding for aa 68 to 206 of the ENase was employed as a single stranded template. We simultaneously added all three mutagenic primers 5'-GCCATTAGATCTTGATCTCCTC, 5'-GACCCCTCTAGAAAAAGGACGT, 5'-TTAACAACGCGTCCATCTGGTC, plus the M13 'universal' primer. These oligos introduced silent mutations creating *Bgl*II, *Xba*I and *Mlu*I sites, at positions corresponding to aa 135, 169 and 187, respectively. A clone scoring positive for all three mutations was chosen and, after sequence verification, used as a source of DNA to return the *Hin*dIII-*Pst*I fragment, bearing the mutations, to the *ecoRIR* gene of pKGS plasmid, substituting for the original region. Combinatorial, site saturation mutagenesis at aa positions 144, 145 and 200 of *Eco*RI endonuclease was approached by the following strategy: M13mp18HPECO, an M13mp18 clone containing the *Hin*dIII-*Pst*I insert of the *ecoRIR* gene was used to infect *E. coli* strain JM101. The ss DNA was purified and used as template for primer extension using the two mutagenic oligos (5'-ATCAAGATCTAATGGCTGCTGGTAATGCTATCXXXXXXTCTCATAAGA, 5'-CGCGTTGTTAATCTTGAGTATAAT-TCTGGTATATTAAATXXXTTAGATCGAC, where the X denote positions mutagenized in the manner described above) in separate experiments. The mutagenic oligos were phosphorylated by T4 polynucleotide kinase using [γ-^{32}P]ATP. The radioactive oligos were annealed to the ss DNA and treated with PolIk for 1 h at 37°C. After inactivating the polymerase by heating at 65°C for 10 min, the DNA was digested with *Mlu*I or *Eco*RI depending on the oligo used in the extension. The ss DNA fragment produced was purified by 20% polyacrylamide-8 M urea gel after visualization of the band by autoradiographic exposure. The ss DNA fragments of 161 and 99 nt (corresponding to the expected lengths of the fragments generated with the oligos complementary to the regions of aa 144–145 and 200, respectively) were gel purified. The fragments (approx. 0.03 pmol each) were then ligated to one another, aided by 0.3 pmol of an adaptor oligo (5'-TTAACAACGCGTCCATCTGGTC) in a 20 μl reaction. The expected 260 nt ss DNA product was gel purified and subsequently amplified by PCR. We used 5'-ATCAAGATCTAATGGCTGCTGGTAA and 5'-CTAGAGTCGACCTGCAGTTA as primers, and performed 30 cycles at 95, 55 and 65°C, for 1.5, 1.5 and 3 min, respectively. The amplified fragment was ligated to adapters (due to the unexpected difficulty to digest with *Bgl*II at the site located 4 bp from the end of the fragment) and then digested with *Bgl*II + *Pst*I. The purified fragment was cloned either in pKGS, replacing the wt with the mutant region as a cassette, or in M13mp19, for direct sequencing.

containing a suitable distribution of triple, double and single replacements, with no stop codons; with methods based on synthesis with mononucleotides, the resultant mutant library would be expected to be poorer, due to inherent limitations imposed by the degeneracy of the genetic code as well as the fact that the triplet is the coding unit. As the advantage of using triplets for mutagenesis, although significant (Sirotkin, 1986), is probably offset by the general difficulty of producing and using such synthetic units, we are currently working at further advancing these methods, adapted to more widely available synthetic chemistries. Our efforts are also stimulated by the versatility attainable by the use of PCR.

(b) **Phenotypes of bacteria bearing the mutant *ecoRIR* genes**

The transformation mixture of the mutant fragment cloned in pKGS was plated in conditions in which the gene is repressed (no IPTG added). The *ecoRIR* gene present in this plasmid carries a mutation that results in a protein product, altered at one aa ($E^{160} \rightarrow D$) (Kuhn et al., 1986), and with lower activity under repressing conditions (note that all of the mutants reported in the present study contain this replacement as well). The presence of this variant of the *ecoRIR* gene, under the repressed conditions, is tolerated but confers a translucent, flat appearance to the *E. coli* JM101 colony bearing it. With the mutant collection, we observed colonies with at least three distinctly different phenotypes: normal, intermediate, and wholly flat-translucent (Fig. 2). Furthermore, when these colonies were replica-plated to media containing the inducer, we could again observe differences among them, namely: normal growth, poor growth and no growth (see Fig. 2 and Table I). To verify the correspondence between the sequence and the observed phenotype, plasmid DNA was purified and used both for retransformation and as a source for the sequencing.

Out of the limited number of colonies analyzed so far (about 10^3, from several transformation experiments), some preliminary observations could be noted. Some 10% of the colonies displayed phenotypes indistinguishable from those carrying the parent gene, under repressing conditions. Sequencing experiments from several of such colonies revealed the presence of the expected silent mutations (coming from codons present in the trinucleotide mixture used to synthesize the oligo, which are synonymous, but different from the original).

Less than 1% of the colonies showed the intermediate phenotype; sequencing of several such clones revealed only the Cys^{144} and Cys^{200} variants. From a few hundred colonies that were replica plated to inducing conditions, about 2% manifested a distinguishably poorer growth ($E^{144} \rightarrow V$, $E^{144} \rightarrow T$, $R^{145} \rightarrow K$, and $E^{144}R^{145} \rightarrow VT$ were identified this way). Overall, therefore, there is an indication that less than 5% of the variant *Eco*RI altered in one or more of these three residues confers a phenotype to the cells carrying them (as detected in our assay system).

Sensitivity to the presence of the M · *Eco*RI was tested for some of the mutants either by transforming the corresponding plasmid DNA into JM101 cells bearing the *ecoRIM* gene in a compatible plasmid or by replacing the mutant fragment in a plasmid that contains both the *ecoRIR*

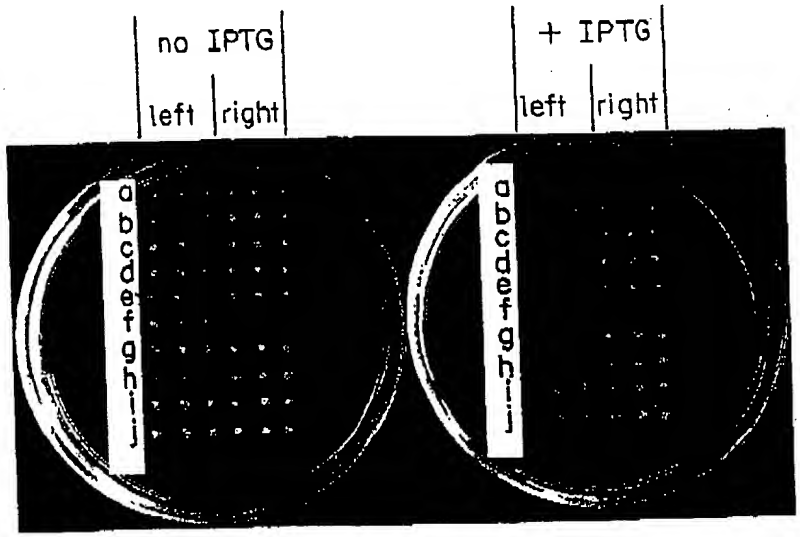

Fig. 2. Colony phenotypes of variant ENases. Colonies expressing *Eco*RI mutant ENases are as follows: a, parent enzyme; b, $E^{144} \rightarrow C$; c, $E^{144} \rightarrow T$; d, $E^{144} \rightarrow G$; e, $E^{144}R^{145} \rightarrow VT$; f, $E^{144}R^{145} \rightarrow VT$ + M · *Eco*RI (present in *cis*, see Table I); g, $R^{145} \rightarrow K$; h, $R^{200} \rightarrow C$; i, $R^{200} \rightarrow K$; and j, $R^{200} \rightarrow N$. Left plate, no IPTG inducer, right plate, 1 mM IPTG. Each plate is divided in two columns containing three independently isolated colonies from a fresh transformation. Colonies in the left and right columns come from host strains lacking and containing the M · *Eco*RI, respectively.

and the *ecoRIM* genes. All mutants tested showed a MTase-sensitive phenotype in one or more conditions (Fig. 2). It was striking to observe variants with remanent activity that are still MTase-sensitive, especially those containing entirely non-conservative (e.g., $E^{144} \rightarrow V$) as well as double ($E^{144} \rightarrow V + R^{145} \rightarrow T$) aa replacements.

Our results are in general agreement with recent reports (Heitman and Model, 1990; Needels et al., 1989; Alves et al., 1989) showing that some replacements at aa 144, 145 and 200 of *Eco*RI generate ENases with reduced activities, but are still specific towards the GAATTC sequence.

The use of combinatorial mutagenesis, together with our assay system and, possibly, the utilization of the mutant at position 160, permitted the identification of the double replacement variant that retains activity against the canonical site. It is noteworthy that the phenotypes conferred by some of the mutations we isolated differ from those previously reported. For instance, mutants $E^{144} \rightarrow T$ and $E^{144} \rightarrow V$ have detectable in vivo activities in our hands, whereas Heitman and Model (1990) listed them as inactive. On the other hand, we observed a null phenotype for $E^{144} \rightarrow G$, listed as marginally active by the same authors. Since our detection system is different and uses a different strain from those previously reported, it is possible that mutations that result in mechanistically different activities (e.g., nicking, dissociation prior to hydrolysis of the second strand, etc.) stand out in one assay, but not in the other. We are currently working on the isolation of the mutations at aa 144, 145 and 200 from the lesion at position 160, which is also a conceivable cause of the discrepancies noted above.

Due to the amendments on the crystal structure of *Eco*RI, there has been caution in interpreting recent studies of *Eco*RI in structural terms (Lesser et al., 1990). Based on the description from the article reporting the revised, refined, coordinate set (Kim et al., 1990), as well as on our own observations using the available α-carbon coordinates (Brookhaven PDB entry 1R1E), we infer that the side chains of aa 145 and 200 are within hydrogen bonding distance of the bases at the major groove of the DNA molecule, with aa 144 also nearby. Therefore, although our results substantiate the notion that additional contacts are involved in the determination of specificity, they clearly pertain to the protein-DNA interface.

(c) Conclusions

In summary, we have implemented a method to combinatorially mutagenize three residues implicated in the specific recognition of DNA by endonuclease *Eco*RI (McClarin et al., 1986). Our collection is composed of a balanced proportion of single, double and triple replacements and allowed us to isolate several mutants that retain activity. The observation of an increasing number of altered proteins with replacements at these three residues, that still retain activity and specificity, reinforces the experimental support of the notion that other aa play a significant role in the recognition and cleavage processes. Work under way is aimed at the screening of a sufficiently large number of clones in order to score the phenotypes of other multiple replacements as well as extending the mutagenesis window. We believe that our approach, applied to all residues involved in hydrogen bonding interactions, as interpreted from the new structure, should be valuable in providing information with regard to the possible overdetermination of the specificity of this enzyme.

ACKNOWLEDGEMENTS

We are grateful to Paul Gaytán for oligo synthesis, Maria Elena Munguia for technical assistance, Sonia Caro for secretarial help and Mario Zurita for critically reading the manuscript. J.O. is supported by a Ph.D. scholarship from Consejo Nacional de Ciencia y Tecnologia, México.

REFERENCES

Adams, S.P., Kauka, K.S., Wykes, E.J., Holder, S.B. and Gallupi, G.R.: Hindered dialky amino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. J. Am. Chem. Soc. 105 (1983) 661–663.

Alves, J. Rüter, T., Geiger, R., Fliess, A., Manss, G. and Pingoud, A.: Changing the hydrogen-bonding potential in the DNA binding site of *Eco*RI by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site -GAATTC-. Biochemistry 28 (1989) 2678–2684.

Beaucage, S.L. and Caruthers, M.H.: Deoxynucleoside phosphoroamidites. A new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. 22 (1981) 1859–1862.

Betlach, M., Hershfield, V., Chow, L., Brown, W., Goodman, H.M. and Boyer, H.W.: A restriction endonuclease analysis of the bacterial plasmid controlling the *Eco*RI restriction and modification of DNA. Fed. Proc. 35 (1976) 2037–2043.

Broka, C., Hozumi, T., Arentzen, R. and Itakura, K.: Simplifications in the synthesis of short oligonucleotide blocks. Nucleic Acids Res. 8 (1980) 5461–5473.

Dunn, I.S., Cowan, R. and Jennings, P.A.: Improved peptide function from random mutagenesis over short 'windows'. Prot. Eng. 2(4) (1988) 283–291.

Heitman, J. and Model, P.: Substrate recognition by the *Eco*RI endonuclease. Proteins 7 (1990) 185–197.

Ito, H., Ike, Y., Ikuta, S. and Itakura, K.: Solid phase synthesis of polynucleotides, VI. Further studies on polystyrene copolymers for the solid support. Nucleic Acids Res. 10 (1982) 1755–1769.

Kim, Y., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M.: Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. Science 249 (1990) 1307–1309.

Kuhn, I., Stephenson, F.H., Boyer, H.W. and Greene, P.J.: Positive-selection vectors utilizing lethality of the *Eco*RI endonuclease. Gene 44 (1986) 253–263.

Lesser, D.R., Kurpiewski, M.R. and Jen-Jacobson, L.: The energetic basis of specificity in the *Eco*RI endonuclease-DNA interaction. Science 250 (1990) 776–786.

McClarin, J.A., Frederick, C.A., Wang, B., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M.: Structure of the DNA-*Eco*RI endonuclease recognition complex at 3 Å resolution. Science 234 (1986) 1526–1541.

Messing, J.: A multipurpose cloning system based on the single stranded DNA bacteriophage M13. Recombinant DNA Technical Bulletin, NIH Publication No. 79-99 2(2) (1979) 43–48.

Needels, M.C., Fried, S.R., Love, R., Rosenberg, J.M., Boyer, H.W. and Greene, P.J.: Determinants of *Eco*RI endonuclease sequence discrimination. Proc. Natl. Acad. Sci. USA 86 (1989) 3579–3583.

Oelgeschläger, T., Geiger, R., Rüter, T., Alves, J., Fliess, A. and Pingoud, A.: Probing the function of individual amino acid residues in the DNA binding site of the *Eco*RI restriction endonuclease by analysing the toxicity of genetically engineered mutants. Gene 89 (1990) 19–27.

Reidhaar-Olson, J.F. and Sauer, R.T.: Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241 (1988) 53–57.

Sartorius, J., Lehming, N., Kisters, B., Von Wilcken-Bergmann, B. and Müller-Hill, B.: Lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. EMBO J. 8(4) (1989) 1265–1270.

Sirotkin, K.: Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. J. Theor. Biol. 123 (1986) 261–279.

Su, T.-Z. and El-Gewely, M.R.: A multisite-directed mutagenesis using T DNA polymerase: application for reconstructing a mammalian gene. Gene 69 (1988) 81–89.

Wells, J.A.: Additivity of mutational effects in proteins. Biochemistry 29 (1990) 8509–8517.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 (1985) 103–119.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 266, No. 13, Issue of May 5, pp. 8495-8500, 1991
Printed in U.S.A.

Combinatorial Mutagenesis of the Reactive Site Region in Plasminogen Activator Inhibitor I*

(Received for publication, September 6, 1990)

John D. York‡, Ping Li, and Stephen J. Gardell§

From the Department of Biological Chemistry, Merck Sharp & Dohme Research Laboratories, West Point, Pennsylvania 19486

Plasminogen activator inhibitor (PAI-1) rapidly inactivates tissue plasminogen activator (t-PA) and urokinase (UK) with nearly identical association rate constants. The contributions of Ser^{344}, Ala^{345}, and Arg^{346} (P_3, P_2, and P_1 residues, respectively) in PAI-1 to inhibition of UK and t-PA were evaluated using combinatorial mutagenesis of the human PAI-1 cDNA. A bacteriophage λ expression library potentially encoding the 8000 unique PAI-1 species were screened for inhibitory activity against UK using a fibrin indicator gel. 390 plaques demarcated by zones of retarded fibrinolysis were analyzed to determine the DNA sequences of their associated active PAI-1 species. We found 134 unique PAI-1 variants that retained inhibitory activity towards UK; they contained a variety of amino acids in their P_3 and P_2 positions but only Arg or, infrequently, Lys in their P_1 position. Each of the unique active PAI-1 were assayed for inhibitory activity towards UK or t-PA; many substitutions differentially affected the ability of the inhibitor to inactivate UK and t-PA. For example, replacement of Ser^{344} and Ala^{345} with Val and Pro, respectively, yielded a PAI-1 variant exhibiting an association rate constant that was unchanged for t-PA but decreased 23-fold for UK, relative to native PAI-1. In general, the PAI-1 variants were more potent inhibitors of t-PA than UK. Hence, t-PA appears more tolerant than UK of structural diversity present in the P_3 and P_2 positions of the PAI-1 variants.

The amino acid sequence of human PAI-1,[1] as deduced from its cDNA sequence (1–3), reveals that PAI-1 is a member of the serpin class of serine proteinase inhibitors (4, 5). Serpins form a complex with their target proteinase that is presumed to involve an intermolecular covalent bond. Formation of this putative covalent linkage would result in proteolytic cleavage of the inhibitor. NH_2-terminal sequencing of the COOH-terminal proteolytic fragment of PAI-1 generated from its interaction with t-PA confirms that the P_1 and P_1' positions[2] of the reactive site of PAI-1 are occupied by Arg^{346} and Met^{347}, respectively (6).

The association of proteinase inhibitors with their target proteinases typically involves amino acid residues in the inhibitor in addition to those situated at the reactive site, P_1 and P_1' (7, 8). These secondary interactions contribute to the exquisite specificity and efficacy which is exhibited by many proteinase inhibitors. Amino acid residues in the inhibitor that are contiguous to the reactive site frequently occupy subsites in the extended active site region of the proteinase and presumably are key determinants of the interaction. Studies using amidolytic substrates suggest that the extended active site regions of t-PA and UK contain S_3 and S_2 subsites (9). We hypothesized that the P_3 and P_2 residues of PAI-1 occupy these subsites and thereby contribute to the stability of the inhibitor-proteinase complex.

The random and simultaneous substitution of amino acids in multiple positions of a protein sequence by combinatorial casette mutagenesis has proven to be a powerful strategy for generating a large number of variant proteins that can subsequently be monitoried for consequences on activity (10–12). We have applied this approach to PAI-1 and used it to evaluate the contribution of the P_3 through P_1 residues of PAI-1 to inhibitory activity. Knowledge of the amino acid substitutions in PAI-1 that are tolerated without abolishing inhibitory activity provides insight into the specificity restraints imposed by the active site regions of these plasminogen activators.

EXPERIMENTAL PROCEDURES

Materials—Human plasminogen (Glu-type), Desafib, Spectrozyme P1, PAI-1 monoclonal antibody 379, and ImmunoBind PAI-1 ELISA kits were from American Diagnostica (Greenwich, CT). Recombinant t-PA, predominantly single-chain, was from Genentech (South San Francisco, CA); it was converted to two-chain t-PA by treatment with plasmin-Sepharose as described elsewhere (13). Human thrombin was from Sigma. Bovine fibrinogen was from CalBiochem (LaJolla, CA). Restriction enzymes were from New England Biolabs (Beverly, MA). T4 DNA ligase was from Boehringer Mannheim. Native Taq polymerase was from Perkin-Elmer Cetus (Norwalk, CT). Lambda Zap II, *Escherichia coli* strain XL-1 Blue, and GigaPack Gold were from Stratagene (LaJolla, CA). Metal chelate-Sepharose and NAP-10 gel filtration columns were from Pharmacia LKB Biotechnology Inc. [α-^{35}S]dATP was from Amersham Corp. Centriprep-30 devices were from Amicon (Danvers, MA). Sequenase Version 2.0 sequencing kits were from U. S. Biochemical (Cleveland, OH). Pyr-Gly-Arg-4-methylcoumaryl-7-amide and MeoSuc-Ile-Gly-Arg-7-amido-4-methylcoumarin were from Peninsula Laboratories (Belmont, CA) and Enzyme Systems Products (Livermore, CA), respectively. The chromogenic substrate S-2288 was from Kabi-Vitrum (Stockholm, Sweden). Oligonucleotides were synthesized with an Applied Biosystems

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present Address: Div. of Hematology-Oncology, Washington University School of Medicine, St. Louis, MO 63110.

§ To whom correspondence should be addressed.

[1] The abbreviations used are: PAI-1, plasminogen activator inhibitor type I; t-PA, two-chain tissue plasminogen activator; UK, urokinase; bp, base pair(s); LB, Luria broth; Hepes, 4-(2-hydroxyethyl)-1-piperazinoethanesulfonic acid; kb, kilobase(s); ELISA, enzyme-linked immunosorbent assay.

[2] The terminology for the amino acid residues of the inhibitor in the vicinity of the reactive site (P_n and P_n') and the complementary subsites of the plasminogen activators (S_n and S_n') are adapted from Schechter and Berger (25).

380B DNA synthesizer (Applied Biosystems, Foster City, CA).

Construction of PAI-1 Variant Expression Library—A degenerate double-strand DNA cassette spanning the coding region for Gln^{322}-Pro^{379} of human PAI-1 and encoding all possible amino acid combinations in the P_3, P_2, and P_1 positions (amino acids 344, 345, and 346, respectively) was synthesized using the following four oligonucleotides (where *N* is a mixture of G, A, T, and C, and *M* is a mixture of A and C): (*a*) 5′GCGCAGGCGCTGCAGAAAGTGAAGATCGA GGTGAACGAGAGTGGTACCGTGGCTAGCTCATCCACG3′, (*b*) 5′GATGATCTCCTCGGGAGCCATMNNMNNMNNGACTATGA CGGCCGTGGATGAGCTAGCCACGGTACC3′, (*c*) 5′CGCATGGC TCCCGAGGAGATCATCATGGACAGACCCTTCCTCTTTGTGG TCCGGCACAACCCGACAGGAACAGTC3′, and (*d*) 5′CTGACTG AATTCGCGGCCGCGTTATTACGGCTCCATCACTTGGCCCAT GAAAAGGACTGTTCCTGTCGGGTTGTGCCG3′. Complementary oligonucleotides a and b, and c and d were annealed by heating to 94 °C for 5 min followed by slow cooling to 37 °C over 1 h. The annealed pairs of oligonucleotides were filled-in using Taq polymerase and dNTPs (14). These double-stranded DNA fragments were digested with *Ava*I, joined using T4 DNA ligase, and digested with *Pst*I, thereby generating the 203-bp mutant cassette (14). Digestion of the human PAI-1 cDNA (courtesy of Dr. J. Han) with *Apa*LI and *Pst*I released a fragment encoding His^2 through Leu^{321}. This fragment was joined to a synthetic tryptophan promoter/operator region flanked by 5′-*Eco*RI and 3′-*Nde*I sites (courtesy of Dr. G. Vlasuk) using an oligonucleotide linker with 5′-*Nde*I and 3′-*Apa*LI ends that contribute the codons for the initiation Met and Val^1 of the PAI-1 amino acid sequence. DNA sequencing using the dideoxy chain termination method (15) was performed to verify the integrity of the promoter/cDNA junction. The promoter/partial cDNA fragment was ligated to the 203-bp mutant cassette and digested with *Eco*RI and *Not*I to yield a 1.3-kb promoter/mutant cDNA expression cassette. This cassette was then inserted into lambda Zap II and packaged using GigaPack Gold to yield a total of 1.2×10^6 plaque-forming units. The expression library contained 40% recombinants as determinants by PAI-1-specific nucleic acid hybridization. Generation of the PAI-1 variant expression library required the use of several restriction endonucleases that recognize nascent sites in the degenerate region of the cDNAs and abolished 384 out of the 32,768 possible cDNA sequences. None of the predicted cleavages eliminated a unique amino acid variant.

Functional Screening of Active PAI-1 Variants—The bacteriophage PAI-1 variant expression library (80,000 recombinants) was incubated with *E. coli* strain XL-1 Blue, diluted with LB top agarose (0.7%), poured onto LB agar (1.2%) plates (150-mm diameter), and grown for 10 h at 37 °C. The resultant plaques were monitored for the presence of PAI-1 activity by a modification of reverse fibrin autography (16) as follows. The plaques were placed in contact with spontaneously lysing indicator gels containing human fibrinogen (2.1 mg/ml), human Glu-plasminogen (6 μg/ml), human thrombin (60 mIU/ml), human UK (12 pM), and agarose (1%). The agar plates were removed after 1.5 h at 37 °C, and the indicator gels were further developed for 6 h at 25 °C. Positive λ clones were identified, eluted, replated, and subjected to a second round of functional screening. Double-stranded plasmids were excised from isolated bacteriophage using the Stratagene protocol and the nucleotide sequences of the entire synthetic regions of the PAI-1 cDNAs were determined.

Analysis of Variant PAI-1 Species—Cultures of *E. coli* strain XL-1 Blue transformed with plasmids containing promoter/PAI-1 mutant cDNAs were grown to saturation in 5 ml of LB containing ampicillin (50 μg/ml). Cells were harvested, resuspended in 100 μl of 2 M guanidine HCl, 0.1 M sodium citrate, pH 7.2, sonicated 30 s on ice, and centrifuged 5000 × *g* for 10 min. PAI-1 antigen levels in the soluble fractions were quantified by PAI-1-specific ELISA that was calibrated with recombinant yeast-derived native PAI-1 (17). Soluble fractions were adjusted to 4.0 M guanidine HCl and 475 nM PAI-1 antigen. Following incubation at 37 °C for 1 h, the lysates were diluted 400-fold and further incubated with 100 pM UK or t-PA for 15 min at 37 °C in a volume of 100 μl. Subsequently 0.4 mM Spectrozyme PL, 80 μg/ml human Glu-plasminogen (and 100 μg/ml Desafib for t-PA assays) was added to yield a 200-μl final volume. A_{405} values were continuously monitored for 1 h at 25 °C using a V_{max} microplate reader (Molecular Devices, Palo Alto, CA). Residual plasminogen activator activity was determined by comparing velocities (A_{405}/time) to UK or t-PA standard curves.

Purification of Selected Recombinant PAI-1 Variants—Cultures (200 ml) of *E. coli* strain XL-1 Blue transformed with plasmids containing the promoter/PAI-1 mutant cDNAs were grown to saturation in LB containing ampicillin. Cells were harvested, resuspended in 50 ml of 6 mM sodium phosphate, 150 mM NaCl, pH 7.2 (phosphate-buffered saline), and broken by passage through a Stansted cell disrupter (Stansted Fluid Power, Stansted, United Kingdom). The cell homogenate was clarified by centrifugation at 100,000 × *g* for 40 min and the supernatant fraction was diluted with an equal volume of 50 mM Hepes, 0.2 M NaCl, 0.01% Tween 80, pH 7.4. The diluted supernatant was applied to a metal chelate-Sepharose column charged with $ZnCl_2$ and equilibrated with 20 mM Hepes, 0.25 M NaCl, 0.01% Tween 80, pH 7.5. The column was washed exhaustively with the equilibration buffer and eluted with a 0–0.2 M imidazole gradient. PAI-1-containing fractions were identified by immunoblotting with a PAI-1-specific monoclonal antibody, pooled, and concentrated using a Centriprep-30 device. The recovered PAI-1 species were typically greater than 25% pure as judged by comparison of the values determined with the PAI-1-specific ELISA and the micro bicinchoninic acid protein assay (Pierce Chemical Co.).

Determination of Association Rate Constants—The purified recombinant PAI-1 variants were activated by treatment with 4.0 M guanidine HCl at 37 °C (18) for 1 h followed by desalting using NAP-10 gel filtration columns. The concentrations of the activated PAI-1 species were determined by quantitative neutralization of t-PA or UK in a direct amidolytic assay using S-2288 (17). The kinetics of inhibition were monitored as follows: plasminogen activators (1 nM) were added to phosphate-buffered saline containing activated PAI-1 (2–10 nM) and a fluorogenic substrate (either MeoSuc-Ile-Gly-Arg-7-amido-4-methylcoumarin or Pyr-Gly-Arg-4-methylcoumaryl-7-amide). The Michaelis-Menten constants of t-PA for MeoSuc-Ile-Gly-Arg-7-amido-4-methylcoumarin and UK for Pyr-Gly-Arg-4-methylcoumaryl-7-amide were 0.63 and 0.078 mM, respectively. The fluorescence was monitored with a Perkin-Elmer LS-3 fluorescence spectrophotometer at excitation and emission wavelengths of 380 and 460 nM, respectively. The time-dependent inhibition of plasminogen activators at varying concentrations of PAI-1 was recorded as a family of progress curves (19). The data from each curve were fit to the integrated first-order rate equation: $F = v_s t + (v_o - v_s)(1 - e^{-kt})/k$ by nonlinear regression, which allowed for the calculation of the apparent constant k where $k = k_2 + [k_1/(1 + s/K_m)]I$. A plot of k as a function of PAI-1 concentration (I) yielded a line whose slope was equal to $k_1/(1 + s/K_m)$, from which the second-order rate constant, k_1 or k_{assoc}, was calculated.

RESULTS

A set of variant PAI-1 cDNAs encoding all possible amino acid combinations in the P_3, P_2, and P_1 positions was constructed by cassette mutagenesis using synthetic degenerate oligonucleotides. The modified PAI-1 cDNAs were joined to a tryptophan promoter/operator region and inserted into bacteriophage λ to generate a PAI-1 variant expression library (Fig. 1). PAI-1 variants were screened for inhibitory activity against UK using a spontaneously lysing fibrin indicator gel

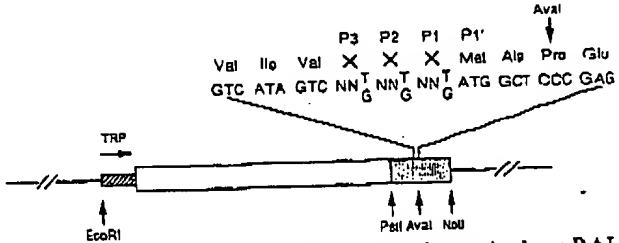

FIG. 1. **Schematic representation of the bacteriophage PAI-1 variant expression library.** The set of degenerate PAI-1 cDNAs was assembled by joining the 5′ end of the PAI-1 cDNA (*open region*) with synthetic oligonucleotides (*stippled region*) designed from the 3′ end of the PAI-1 cDNA. The nucleotide and predicted amino acid sequences in the vicinity of the reactive site of PAI-1 are shown. Positions designated by *N* represent equal mixtures of G, A, T, and C used during oligonucleotide synthesis. The amino acids occupying the P_3, P_2, and P_1 positions are designated by *X* reflecting random replacements in this region. The tryptophan promoter/operator is designated *TRP* and depicted as the *cross-hatched region*. The λ arms are represented by the *flanking lines*. Relevant restriction endonuclease sites are indicated.

containing plasminogen and UK. Bacteriophage expressing an active PAI-1 species were identified by opaque zones of retarded fibrinolysis in the indicator gel. At least 500 zones were readily detected from a sampling of 80,000 recombinant bacteriophages.

390 of the bacteriophages encoding an active PAI-1 species and 42 randomly selected inactive PAI-1 species were isolated, the double-stranded plasmids harboring the promoter/PAI-1 variant cDNAs were excised, and the nucleotide sequences of each cDNA in the region of the P_3, P_2, and P_1 residues were determined. Nucleotide frequencies were calculated from the nine bases encoding the P_3–P_1 positions of the randomly selected inactive PAI-1 cDNA variants. The observed frequencies of G, A, T, and C equalled 0.28, 0.17, 0.32, and 0.23, respectively. These values were very similar to the predicted frequencies of G, A, T, and C that equalled 0.30, 0.20, 0.30, and 0.20, respectively.

The PAI-1 variant cDNAs that were analyzed encoded 135 unique active and 42 unique inactive PAI-1 species (Table I). 121 of the 135 active PAI-1 species retained an Arg in the P_1 position, while the remainder contained a Lys substitution in this position. 14 of 19 and 16 of 19 possible PAI-1 variants containing sole amino acid substitutions in either the P_3 or P_2 positions, respectively, were active towards UK. 90 of the active PAI-1 variants contained simultaneous substitutions in the P_3 and P_2 positions; this value corresponds to 25% of the theoretical maximum. Only seven PAI-1 variants containing simultaneous substitutions in the P_3, P_2, and P_1 positions were identified as being active towards UK; each of these contained Lys in the P_1 position. Primary structures of the 42 randomly picked inactive PAI-1 variants possessed an apparent random distribution of amino acids in the P_3, P_2, and P_1 positions.

The 135 active and 42 randomly picked inactive PAI-1 species were individually synthesized in *E. coli* using the rescued PAI-1 expression plasmids. Immunoblot analysis of bacterial lysates fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (20) (Fig. 2) demonstrated that the recombinant PAI-1 synthesized in *E. coli* (*lane 1*) comigrated with PAI-1 internally expressed in yeast (*lane 3*) (17). The electrophoretic mobilities of each of 12 different active and inactive PAI-1 variants were virtually identical to *E. coli*-derived native inhibitor (data not shown). The immunoreactive species was not present in *E. coli* lysates harboring the parent expression vector but lacking the PAI-1 cDNA (*lane 2*). PAI-1 purified from HT1080 fibrosarcoma cells (*lane 4*) exhibited a slightly larger apparent M_r value due to the presence of carbohydrate, a characteristic not shared by *E. coli* and yeast-derived recombinant PAI-1.

Guanidine-treated lysates containing equal amounts of PAI-1, as judged by PAI-1-specific ELISA, were incubated with UK for 15 min at 37 °C and the residual plasminogen activator activities were assayed. The PAI-1 variants that were sorted with respect to position(s) of the amino acid substitution(s) (Table I) are ranked within these groups according to their abilities to inhibit UK. None of the 42 randomly picked inactive variants exhibited measurable activity toward UK. Only 35 of the 134 active PAI-1 variants exhibited greater than 60% of the inhibitory activity against UK relative to that of native PAI-1. 20 of these 35 contained sole substitutions in the P_3 or P_2 position; 14 of the 35 contained simultaneous substitutions in the P_3 and P_2 positions; and only one of the 35 contained concurrent replacements at all three positions. Ala, Gly, Thr, or Ser was present at the P_3 and P_2 positions in 25 and 22 instances, respectively, of the 35 PAI-1 variants exhibiting greater than 60% relative inhibitory activity toward UK. Ser or Gly was present at the P_2 position in the seven most active PAI-I variants containing substitutions in both the P_3 and P_2 positions.

The 135 PAI-1 species that were active towards UK were also monitored for their abilities to inhibit t-PA. The data, shown in Table I, clearly indicate that many of the PAI-1 variants are differentially active against UK and t-PA. Furthermore, 84 PAI-1 variants retained greater than 60% of the inhibitory activity of native PAI-1 toward t-PA; this number is 2.4-fold higher than that observed for UK. A similar functional screen of the PAI-1 variant expression library using t-PA instead of UK yielded approximately four times more zones of retarded fibrinolysis (data not shown).

Recombinant native PAI-1 and selected PAI-1 variants were partially purified from bacterial cell lysates using Zn-chelate-Sepharose chromatography. The PAI-1 species were activated with guanidine and their functional molarities were determined by titration against known amounts of UK or t-PA. The apparent k_{assoc} values of these PAI-1 species, HT1080 PAI-1 and yeast-derived recombinant PAI-1, were determined with both UK and t-PA using a continuous fluorometric assay (Table II). The apparent k_{assoc} constants exhibited by the *E. coli*-derived native PAI-1 for t-PA and UK are similar to one another. Likewise, they are similar to those determined with yeast-derived native PAI-1 but slightly less than those obtained with HT1080 PAI-1. This difference between yeast-derived PAI-1 and HT1080 PAI-1 was previously noted (17) and is probably due to the absence of carbohydrate on the inhibitor expressed in yeast.

The apparent k_{assoc} constants obtained with the selected PAI-1 variants (Table II) were consistent with the relative inhibitory activities shown in Table I and confirmed that certain amino acid substitutions in the P_3 and P_2 positions of PAI-1 selectively crippled inhibitory activity against UK or t-PA. For example, replacement of Ala^{346} in PAI-1 with Arg or Lys resulted in larger decreases in the apparent k_{assoc} constants with t-PA than with UK. On the other hand, several of the chosen PAI-1 variants (Tyr^{344}; Val^{344}, Pro^{345}; Val^{344}, Ser^{345}; Tyr^{344}, Ser^{345}; Tyr^{344}, Gly^{345}) exhibited apparent k_{assoc} values that were significantly diminished for UK but essentially unchanged for t-PA. Two of these variants (Tyr^{344}, Ser^{345} and Tyr^{344}, Gly^{345}) displayed an apparent k_{assoc} constant that was somewhat higher than that of the *E. coli*-derived native inhibitor. The most impressive example of altered specificity is the Val^{344}, Pro^{345} double-substitution variant. Its apparent k_{assoc} value with t-PA was similar to that of native PAI-1, but the apparent k_{assoc} value with UK was decreased by approximately 23-fold.

DISCUSSION

Amino acid substitutions in the P_3, P_2, and P_1 positions of PAI-1 can have a profound effect on the ability of the inhibitor to inactivate UK and t-PA. A striking paradigm revealed by this study is that the activity of PAI-1 toward UK is contingent on the presence of Arg or, less frequently, Lys in the P_1 position. Concurrent changes in the neighboring P_3 and P_2 residues do not circumvent the strict necessity for a basic amino acid residue in the P_1 position. This obligatory requirement for Arg or Lys is concordant with the presence of an Arg in the P_1 position of plasminogen (21), the preferred substrate of UK and t-PA.

Amino acid substitutions in the P_3 and P_2 positions of PAI-1 were generally well tolerated by UK, although not all replacements were allowed. 14 of 19 possible P_3 variants, 16 of 19 possible P_2 variants, and 90 of 361 possible P_3 and P_2

TABLE I

Inhibitory activity of the unique PAI-1 species toward UK (column U) and t-PA (column T)

Single-letter code is used to designate the amino acids occupying the P_3, P_2, and P_1 positions of each PAI-1 species. -, consensus with the native PAI-1 sequence; ., a termination codon. The active PAI-1 variants are grouped with respect to position(s) of the substitution(s). Within each of these groups, species are ranked according to inhibitory activities toward UK. The corresponding inhibitory activities toward t-PA are also shown. Inhibitory activity is expressed relative to that exhibited by native PAI-1 (outlined by the box). These values represent the average of four independent assays and in all cases the standard deviation was less than 10%. The inactive PAI-1 variants are grouped separately.

NATIVE PAI-1

	U	T
SAR	100	100

ACTIVE P_3 VARIANTS

T--	115	110
C--	102	98
A--	100	100
M--	92	101
P--	86	101
C--	85	100
H--	80	111
L--	73	103
Y--	70	119
F--	59	102
I--	51	99
N--	41	100
Q--	39	110
E--	37	97

ACTIVE P_2 VARIANTS

-G-	120	92
-Q-	116	74
-S-	109	85
-M-	100	63
-R-	93	24
-H-	85	59
-T-	85	97
-E-	84	61
-K-	82	31
-N-	72	57
-P-	66	100
-C-	56	28
-Y-	45	43
-D-	42	32
-L-	33	73
-V-	30	64

ACTIVE VARIANTS WITH A P_1 LYS

TSK	63	37
T-K	55	59
--K	48	26
-SK	30	10
A-K	25	24
MSK	21	55
ASK	16	17
TRK	16	0
-RK	15	0
-MK	13	7
LPK	13	42
NGK	13	62
-QK	8	0
TTK	8	12

ACTIVE P_3 AND P_2 VARIANTS

	U	T
AS-	107	111
VS-	101	119
TG-	93	97
AG-	91	107
TS-	89	94
MG-	86	101
GG-	85	97
VH-	76	100
TQ-	71	75
AM-	67	83
VG-	66	115
TE-	62	85
TT-	61	77
AK-	61	77
VD-	58	97
MS-	57	107
GS-	56	74
TN-	52	77
LG-	52	108
GQ-	50	58
IR-	46	39
VL-	45	98
LP-	43	92
TK-	43	47
YS-	43	106
LV-	42	94
LS-	42	95
TV-	41	58
MT-	41	99
AR-	40	43
HS-	39	95
AN-	39	83
YH-	38	99
YG-	38	110
YQ-	38	103
FV-	38	95
FG-	37	98
VN-	37	74
VP-	37	119
AT-	37	92
TH-	36	50
TL-	36	87
LL-	35	90
FS-	35	95
TP-	35	104
NG-	35	99
YI-	34	52
MP-	33	106
HM-	32	75
TD-	31	40
HQ-	31	72
RG-	31	100
VT-	30	109
MR-	29	96
FQ-	29	92
GR-	29	14
QS-	28	91
NQ-	25	50
LQ-	25	81
NM-	24	46
CG-	23	64
FG-	23	62
FS-	21	30
AH-	21	55
MR-	21	30
MH-	20	82
LE-	20	84
LR-	20	24
GL-	20	65
QR-	19	22
LT-	18	85
NR-	18	17
CS-	18	53
DL-	18	61
HC-	17	22
CN-	16	24
VH-	15	109
TY-	14	46
MV-	14	66
KR-	13	9
AI-	13	61
QL-	13	84
YR-	11	45
AL-	9	91
GN-	9	10
PH-	9	9
TW-	8	37
HR-	2	29
TF-	1	46
LM-	0.3	60

INACTIVE VARIANTS

AII	PGP	NWI
APP	PSH	NSP
AMA	PQ.	QTD
ASL	SVN	QYP
ADL	SPA	DGV
ARP	SDE	DTH
VSH	SKW	DRP
VQW	SR.	KQP
LLG	SHP	RLN
LID	TPQ	RPF
IRH	CPS	RMG
I.P	YSG	RCP
FPX	NAL	REF
PPS	NPV	R.Q

double variants were identified as active inhibitors of UK by the functional screen.

Active PAI-1 variants containing the reactive site residues, Arg-Met, in the original P_3-P_2 or P_2-P_1 sites were not identified by the functional screen. Hence, there appears to be a strict positional requirement for the P_1 and P_1' residues along the putative reactive site loop (5). In contrast, the reactive site loop structure of another serpin inhibitor, α_2-antiplasmin, contains separate, although overlapping reactive sites for trypsin and chymotrypsin (22).

The amino acids in the P_3 and P_2 positions of the PAI-1 variants that are preferred by UK were better appreciated when the PAI-1 variants were ranked with respect to their relative inhibitory activities. The large number of active PAI-1 variants necessitated that a simple "fixed time point" assay be used to determine inhibitory activity. A possible pitfall of this assay, apart from its relative insensitivity to the inhibition kinetics, stems from the existence of latent and active states of PAI-1 (18). Each variant or native PAI-1 preparation was activated with guanidine prior to assay of inhibitory

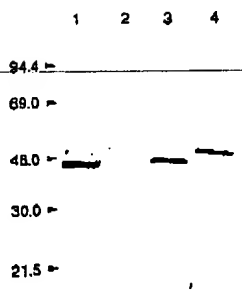

FIG. 2. **Heterologous expression of human PAI-1 in *E. coli*.** Lysates from *E. coli* cultures harboring the expression vector for native PAI-1 (*lane 1*) or the parent vector lacking the PAI-1 cDNA (*lane 2*) were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Purified recombinant PAI-1 synthesized in yeast (120 ng, *lane 3*) and purified PAI-1 secreted from HT1080 fibrosarcoma cells (115 ng, *lane 4*) were also electrophoresed. PAI-1 was detected by immunoblot analysis using an anti-PAI-1 murine monoclonal antibody and alkaline phosphatase-conjugated goat anti-mouse IgG. Protein relative molecular mass markers shown in kiloDaltons are as indicated.

TABLE II

Association rate constants for the interactions between PAI-1 species and t-PA or UK

PAI-1 species: nPAI-1, native PAI-1 secreted from HT1080 fibrosarcoma cells; rPAI-1, recombinant yeast-derived native PAI-1; the *E. coli*-derived PAI-1 species are designated by one-letter amino acid code triplets corresponding to the identity of the amino acids in the P_3, P_2, and P_1 positions. Association rate constants with UK or t-PA were determined as described under "Experimental Procedures."

PAI-1 species	t-PA	UK	t-PA/UK
	$M^{-1}s^{-1}$		
nPAI-1	$(9.2 \pm 3.1) \times 10^6$	$(11.7 \pm 0.8) \times 10^6$	0.8
rPAI-1	$(6.2 \pm 0.5) \times 10^6$	$(6.4 \pm 0.7) \times 10^6$	1.0
SAR	$(7.5 \pm 0.8) \times 10^6$	$(7.4 \pm 0.4) \times 10^6$	1.0
SRR	$(5.4 \pm 0.3) \times 10^6$	$(2.0 \pm 0.2) \times 10^6$	0.3
SKR	$(7.6 \pm 0.1) \times 10^6$	$(1.3 \pm 0.3) \times 10^6$	0.6
VPR	$(7.0 \pm 3.1) \times 10^6$	$(3.1 \pm 0.3) \times 10^6$	23
VSR	$(7.0 \pm 3.1) \times 10^6$	$(1.7 \pm 0.1) \times 10^6$	4
YAR	$(6.3 \pm 1.7) \times 10^6$	$(1.9 \pm 0.1) \times 10^6$	3.3
YSR	$(12.1 \pm 1.5) \times 10^6$	$(1.3 \pm 0.1) \times 10^6$	9.3
YGR	$(10.7 \pm 2.0) \times 10^6$	$(1.6 \pm 0.1) \times 10^6$	6.7

activity. Hence, the potency differences shown in Table I could reflect large variations in the relative amounts of active and latent forms of the inhibitor among the activated PAI-1 preparations. However, the apparent rate constants of selected PAI-1 variants (Table II), which were based on careful titration of active inhibitor, mirror the relative inhibitory activities displayed in Table I. We conclude that disparate amounts of active and latent forms of PAI-1 was not primarily responsible for the broad spectrum of inhibitory activities exhibited by this set of PAI-1 variants.

Certain inferences can be drawn from our data with respect to the topology of the active site region in UK. As stated previously, the S_1 subsite of UK (as well as t-PA) displayed a fastidious requirement for the presence of a basic amino acid residue in the P_1 position of the inhibitor. Furthermore, the 35 most-active PAI-1 variants against UK contained Ala, Gly, Thr, or Ser at the P_3 and P_2 positions in 25 and 22 instances, respectively. These four amino acids are contiguously and centrally located on an empirical hydrophobicity scale derived from the average area that each amino acid buries upon folding in globular proteins of known structure (23). Hence, the S_3 and S_2 subsites of UK apparently prefer amino acids that fall within a narrow range of intermediate hydrophobicity.

Comparison of the inhibitory activities of the PAI-1 variants against UK and t-PA reveals that t-PA is more tolerant than UK of amino acid substitutions in the reactive site region of PAI-1. Of the 134 active PAI-1 variants, 35 and 84 exhibited greater than 60% of the inhibitory activity of native PAI-1 toward UK and t-PA, respectively. The apparent topological dissimilarities between t-PA and UK in their active site regions are convincingly shown by the marked differential inhibition of these two plasminogen activators by the PAI-1 variants. Without exception, the single-substitution P_3 variants that were less active than native PAI-1 toward UK did not exhibit diminished activities against t-PA. In contrast to UK, t-PA was similarly tolerant of variation in the P_3 position of PAI-1 variants also containing substitutions such as Ser, Gly, or Thr in the P_2 position. We postulate that an S_3 subsite may not exist in t-PA for the binding of PAI-1 or, if it does exist, would readily accommodate a larger variety of amino acid side chains than the S_3 subsite of UK. The ligand preference of the S_2 subsite of t-PA was relatively broad although, unlike UK, there appears to be a bias against basic amino acids.

Assuming that all possible unique PAI-1 cDNAs were equally represented in the expression library, then statistical analysis[3] predicts that greater than 91% of the 32,768 possible PAI-1 cDNAs were present in the sampling of 80,000 recombinant bacteriophages screened with the fibrin indicator gel. Nevertheless, our evaluation of the amino acids tolerated at the reactive site region in PAI-1 is not as complete as predicted from the above analysis for a variety of reasons. First, the large number of positives encountered during the screen precluded analysis of each one. Second, many plaques containing a less-active PAI-1 variant could be missed due to the transitory nature of weaker signals upon continuing development of the indicator gel. Third, bias could be imposed upon the library due to differences in the growth rates or plaque sizes of certain recombinant bacteriophage clones. In spite of these reservations, the presently described screening strategy provided a broad and insightful survey of the consequences of a vast number of amino acid substitutions in the vicinity of the reactive site of PAI-1.

In conclusion, our results have demonstrated the importance of P_3, P_2, and P_1 residues in PAI-1 to inhibitory efficacy towards UK and t-PA. Furthermore, this study highlights dissimilar specificity restraints imposed by the active site regions of these two plasminogen activators. The resourcefulness of the combinatorial mutagenesis approach used herein is clearly evident by the fortuitous, simultaneous amino acid replacements in the P_3 and P_2 positions that gave rise to PAI-1 variants (such as Val^{344}, Pro^{345}) exhibiting a marked preference for t-PA over UK. The PAI-1 variants that preferentially inhibited t-PA are of special interest in light of the ability of PAI-1 to rapidly reverse the bleeding tendency in rabbits following the combined administration of t-PA and aspirin (24). Accordingly, PAI-1 variants that are relatively inert towards UK may be safer antidotes for the treatment of t-PA toxicity. The functional screen described herein will continue to be a valuable tool in probing additional regions in PAI-1 and, after suitable adaptations, this method could provide insight into analogous interactions between other inhibitors and their target proteases.

[3] A 95% confidence upper bound on the number of unique PAI-1 cDNAs represented in the population was obtained using the Chebyshev inequality (28).

Acknowledgments—We thank J. Han for the PAI-1 cDNA, G. Vlasuk for the Trp promoter, C. Bennett for oligonucleotide synthesis, M. J. Zaborowski for computational support, K. Soper for statistical analysis, our colleagues for insightful discussion and critical review, and D. Bonenberger for assisting in the preparation of this manuscript.

REFERENCES

1. Ny, T., Sawdy, M., Lawrence, D., Millan, J. L., and Loskutoff, D. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6776–6780
2. Pannekoek, H., Veerman, H., Lambers, H., Diergaarde, P., Verweij, C., Van Zonneveld, A.-J., and van Mourik, J. A. (1986) *EMBO J.* **5**, 2539–2544
3. Ginsburg, D., Zeheb, R., Yang, A. Y., Rafferty, U. M., Andreasen, P. A., Nielsen, L., Dano, K., Lebo, R. V., and Gelehrter, T. D. (1986) *J. Clin. Invest.* **78**, 1673–1680
4. Carrell, R., and Travis, J. (1985) *Trends Biochem. Sci.* **10**, 20–24
5. Carrell, R., and Boswell, D. R. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds) pp. 403–440, Elsevier, Amsterdam
6. Sanzo, M. A., Marasa, J. C., Wittwer, A. J., Siegel, N. R., Harakas, N. K., and Feder, J. (1987) *Biochemistry* **26**, 7443–7449
7. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593–626
8. Read, R. J., and James, M. N. G. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds) pp. 301–336, Elsevier, Amsterdam
9. White, J., and Elmore, D. T. (1986) *Biochem. Soc. Trans.* **14**, 457–458
10. Reidhaar-Olson, J. F., and Sauer, R. T. (1988) *Science* **241**, 53–57
11. Lim, W., and Sauer, R. T. (1989) *Nature* **339**, 31–36
12. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A., III (1989) *Nature* **340**, 397–400
13. Higgins, D. L., and Vehar, G. A. (1987) *Biochemistry* **26**, 7786–7791
14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
15. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
16. Erickson, L. A., Lawrence, D. A., and Loskutoff, D. J. (1984) *Anal. Biochem.* **137**, 454–463
17. Gardell, S. J., Hare, T. R., Han, J. H., Markus, H. Z., Keech, B. J., Carty, C. E., Ellis, R. W., and Schultz, L. D. (1990) *Arch. Biochem. Biophys.* **2378**, 467–474
18. Hekman, C. M., and Loskutoff, D. J. (1985) *J. Biol. Chem.* **260**, 11581–11587
19. Williams, J. W., and Morrison, J. F. (1979) *Methods Enzymol.* **63**, 437–467
20. Laemmli, U. K. (1970) *Nature* **227**, 680–685
21. Robbins, K. C., Summaria, L., Hsieh, B., and Shah, R. J. (1967) *J. Biol. Chem.* **242**, 2333–2342
22. Potempa, J., Shieh, B.-H., and Travis, J. (1988) *Science* **241**, 699–700
23. Rose G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) *Science* **229**, 834–838
24. Vaughan, D. E., Declerck, P. J., De Mol, M., and Collen, D. (1989) *J. Clin. Invest.* **84**, 586–591
25. Schechter, J., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162
26. Mosteller, F., Rourke, R. E. K., and Thomas, G. B., Jr. (1970) *Probability with Statistical Application*, Ed 2, Addison-Wesley Publishing Company, Reading, MA

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8367–8371, September 1993
Biochemistry

Engineering multiple properties of a protein by combinatorial mutagenesis

(protein engineering/additivity/stability/DNA binding protein)

WARREN S. SANDBERG*† AND THOMAS C. TERWILLIGER‡§

*Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637; and ‡Life Sciences Division, MS M880, Los Alamos National Laboratory, Los Alamos, NM 87545

Communicated by David Eisenberg, June 1, 1993 (received for review May 3, 1993)

ABSTRACT **A method for simultaneously engineering multiple properties of a protein, based on the observed additivity of effects of individual mutations, is presented. We show that, for the gene *V* protein of bacteriophage f1, effects of double mutations on both protein stability and DNA binding affinity are approximately equal to the sums of the effects of the constituent single mutations. This additivity of effects implies that it is possible to deliberately construct mutant proteins optimized for multiple properties by combination of appropriate single mutations chosen from a characterized library.**

One of the long-term goals of the study of mutational effects on protein stability and activity is to devise a method by which mutations can be rationally employed to alter or "engineer" the properties of proteins with predictable results. Recombinant DNA technology has allowed the construction of proteins of altered stability *in vitro* (1–16), catalytic efficiency (17–19), substrate specificity (20–23), and resistance to *in vivo* thermal inactivation (24) through the use of single or multiple amino acid substitutions. This effort has been greatly helped by the fact that the effects of amino acid substitutions on such properties of proteins tend to be additive as mutations accumulate, provided that the substituting residues do not interact functionally or by direct contact (25). For example, additive increases in the stability of subtilisin BPN′ have been achieved by combining mutations at six sites in the protein tertiary structure (16). The six mutations individually stabilize the protein by 0.3–1.3 kcal/mol, and the individual effects sum to a stability increase of 3.8 kcal/mol predicted for the hexa-mutant. The observed stabilization of the mutant containing all six substitutions is 4.3 kcal/mol (16). Additive effects of amino acid substitutions have been used to engineer incremental increases in the stability of other proteins including the N-terminal domain of λ repressor (5, 13), T4 lysozyme (9, 12), kanamycin nucleotidyltransferase (6), and neutral protease (26) as well. This strategy of additive mutation has also been employed to alter binding affinities or specificities of proteins, such as λ repressor (20), subtilisin (21–23), and glutathione reductase (27), for their substrates or cofactors, and to alter the pH profile of subtilisin (17).

A factor complicating the effort to engineer proteins by mutation is that most single amino acid substitutions alter multiple properties of the proteins in which they are made. To be functional, a protein must be at once stable, yet flexible, with high catalytic activity balanced against substrate specificity. Because mutants affecting only one of these properties are relatively rare, it appears difficult to optimize one characteristic of a protein through mutations while maintaining adequacy in the others. However, the observation that mutational effects on the *in vitro* properties of proteins are frequently additive suggests that it may be possible to counteract deleterious side effects of desirable (or primary) mutations by including additional mitigating (or secondary) mutations. If the effects of combining multiple mutations display simple additivity, then the net effect of the primary and secondary mutations on a given *in vitro* property of the subject protein should be the algebraic sum of the effects observed in the starting mutants. The objective of the present study is to test this assumption and to explore the possibility of creating proteins that have been optimized with respect to multiple *in vitro* properties.

The gene *V* protein of bacteriophage f1 is a small single-stranded DNA (ssDNA) binding protein that lends itself to this goal because its DNA binding affinity and stability can be readily measured *in vitro* and mutants of the gene *V* protein are readily available. A plasmid-based mutagenesis and expression system allows the rapid production of proteins containing single and multiple substitutions (28, 29). Conditions for monitoring gene *V* protein stability *in vitro* have been established (30), allowing quantitative assessment of the effects of single and multiple substitutions on stability (3, 4, 31). The stability of wild-type (WT) and mutant gene *V* proteins can be estimated as a function of their resistance to guanidine hydrochloride (Gdn·HCl)-induced denaturation, monitored by the disappearance of tyrosine circular dichroism (CD) at 229 nm as the protein unfolds (30). Cooperative binding of the protein to its substrate, ssDNA, can be followed *in vitro* by monitoring the intrinsic tyrosine fluorescence of the protein (32). This fluorescence is quenched as the protein binds to ssDNA and is restored when the protein–ssDNA complex is dissociated by the addition of NaCl. Binding affinities of WT gene *V* protein to a variety of substrates and of WT and several mutant gene *V* proteins to the substrate polydeoxyadenylic acid have been reported (31, 33–36). Thus the gene *V* protein provides a system in which proteins containing amino acid substitutions may be readily obtained and characterized *in vitro*. We used this system, starting with well-characterized single-substitution mutants, to construct doubly substituted proteins displaying predictable and additive changes in both DNA binding affinity and stability.

MATERIALS AND METHODS

Mutagenesis, Strains, and Vectors. Mutagenesis of gene *V* was carried out in the plasmid pTT18 as described (28, 29). Single mutants were constructed by oligonucleotide-directed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Gdn·HCl, guanidine hydrochloride; ssDNA, single-stranded DNA; WT, wild type. Substitutions are described in the one-letter code; e.g., Y41F denotes the replacement of tyrosine at position 41 by phenylalanine.

†Present address: Pritzker School of Medicine, University of Chicago, Chicago, IL 60637.
§To whom reprint requests should be addressed.

Table 1. Single- and double-substitution mutants of the gene V protein

Mutation(s)	Minimum separation, Å	$\Delta\Delta G^\circ_{u,2M}$, kcal/mol	$\Delta\Delta G^\circ_{d,0.15M}$, kcal/mol
I6V		−0.68	0.21
F13T		−0.67	−0.65
L28V		1.09	−0.06
E30F		1.97	2.29
L32Y		1.04	0.37
C33M		−3.49	−0.45
C33V		−0.18	−0.09
V35A		−2.25	−0.45
V35C		−1.45	−0.62
V35F		−3.21	−1.00
V35I		−0.68	−0.24
V35L		−2.72	−0.47
V35M		−1.10	−0.99
Y41F		−0.62	−3.12
V45C		−0.05	0.15
I47C		−5.30	−0.29
I47F		−2.02	0.85
I47L		−0.67	0.03
I47M		−2.21	0.61
I47V		−2.62	−0.25
H64C		0.50	−0.12
L65P		−1.47	−0.73
F68L		−4.30	−0.26
F73W		0.76	1.21
M77I		1.63	0.10
M77V		1.23	0.06
R82C		−1.50	−2.74
A86T		−0.66	−0.26
A86V		0.47	−0.25
I6V/E30F	10.7	0.71	2.27
I6V/M77I	17.3	0.34	0.17
I6V/M77V	17.3	−0.04	0.27
F13T/E30F	5.8	1.59	1.03
L28V/F68L	19.0	−3.46	−0.84
E30F/A86T	21.6	1.17	1.93
E30F/A86V	21.6	2.05	1.94
L32Y/R82C	15.6	0.15	−1.78
C33V/V35C	4.0	−1.57	−0.57
C33M/I47C	3.9	−5.73	0.29
V35A/I47F	6.4	−3.67	0.14
V35A/I47L	6.4	−2.97	−0.50
V35A/I47M	6.4	−4.51	0.10
V35A/I47V	6.4	−4.52	−0.89
V35C/I47C	6.4	−7.20	−0.79
V35F/I47L	6.4	−4.22	−1.28
V35I/I47F	6.4	−2.08	0.30
V35I/I47L	6.4	−1.18	−0.29
V35I/I47M	6.4	−2.85	0.46
V35I/I47V	6.4	−3.12	−0.51
V35L/I47F	6.4	−4.00	0.24
V35L/I47L	6.4	−3.58	−0.51
V35L/I47M	6.4	−5.41	−0.08
V35L/I47V	6.4	−5.10	−1.03
V35M/I47F	6.4	−2.35	0.07
V35M/I47L	6.4	−1.70	−1.08
V35M/I47M	6.4	−3.62	−0.48
Y41F/F73W	21.0	−0.66	−0.73
V45C/R82C	11.7	−1.05	−2.44
H64C/F68L	7.0	−4.07	−0.80
L65P/F68L	4.5	−4.25	−0.83

For double mutants, the closest pair of atoms (minimum separation) in the two side chains in the crystal structure of the WT gene V protein dimer (M. M. Skinner, H. Zhang, D. H. Leschnitzer, Y. Guan, H. Bellamy, R. M. Sweet, C. W. Gray, R. N. H. Konings, A. H.-J. Wang, and T.C.T., unpublished data) is listed. In two cases, Ile-6/Met-77 and Tyr-41/Phe-73, the two side chains are within separate monomers, and in all other cases, they are within the same monomer of the protein. Changes in stability, measured as the change in free energy upon unfolding ($\Delta G^\circ_{u,2M}$), are given in kcal/mol of dimeric protein, relative to WT. Mutants with increased stabilities have positive values of $\Delta\Delta G^\circ_{u,2M}$ and the magnitude corresponds to making the same substitution twice. Changes in apparent free energies of dissociation from polydeoxyadenylic acid ($\Delta\Delta G^\circ_{d,0.15M}$), relative to WT gene V protein, are given in kcal/mol; positive values of $\Delta\Delta G^\circ_{d,0.15M}$ indicate enhanced binding of the mutant to ssDNA relative to WT.

mutagenesis and isolated as derivatives of pTT18. Double mutants were obtained by oligonucleotide-directed mutagenesis, recombination of single mutants by the use of intervening restriction sites, or selection (as intragenic suppressors of a conditional lethal mutation) from a pool of random single-amino acid substitution mutants (31). Mutant genes were expressed in *Escherichia coli* K561 (37). Transformation of *E. coli* K561 with pTT18 derivatives was effected using an electroporation device.

Protein Purification. Growth of K561 cultures transformed with pTT18 derivatives encoding gene V protein variants and purification of proteins were carried out as described (30, 31). To confirm that the mutant proteins contained the expected amino acid substitutions, ssDNA was isolated from *E. coli* harvested late in the growth and the gene V region was sequenced.

Measurement of ssDNA Binding Affinity. NaCl-induced dissociation of gene V protein–ssDNA complexes, monitored by fluorescence, was used to estimate the binding affinities of WT and mutant gene V proteins for the substrate polydeoxyadenylic acid as described (refs. 31–36 and unpublished observations). Data are reported as the apparent free energy change upon dissociation in 0.15 M NaCl ($\Delta G^\circ_{d,0.15M}$), related to the effective binding constant in 0.15 M NaCl ($K\omega_{0.15M}$) by $\Delta G^\circ_{d,0.15M} = +RT\ln(K\omega_{0.15M})$ (where R = 1.987 cal/mol·K; T = 298 K). Differences in binding between mutants are expressed as differences in free energy change upon dissociation ($\Delta\Delta G^\circ_{d,0.15M}$), defined as [$\Delta G^\circ_{d,0.15M}$ (mutant) − $\Delta G^\circ_{d,0.15M}$ (WT)]. Mutants binding more tightly to ssDNA than the WT will have positive values of $\Delta\Delta G^\circ_{d,0.15M}$. Error estimates (2 SD) were obtained from seven measurements of the DNA binding affinity of the WT gene V protein leading to an error of ±0.1 kcal/mol for $\Delta\Delta G^\circ_{d,0.15M}$.

Measurements of Protein Stability. Stability measurements on mutant gene V proteins were carried out as described (30). The gene V protein is reversibly denatured by Gdn·HCl, and the denaturation can be monitored by the disappearance of a tyrosine CD signal at 229 nm. Unfolding data were fitted to a two-state model (30) with modifications (4) in the case of proteins for which the unfolding is >50% complete when [Gdn·HCl] < 1.5 M. Stabilities are expressed as free energy changes upon unfolding in kcal/mol of dimeric protein. The stability ($\Delta G^\circ_{u,2M}$) of the WT gene V protein, given as the average ± 2 SD of 10 measurements, is 9.04 ± 0.3 kcal/mol of dimeric protein. Stabilities of mutant ($\Delta G^\circ_{u,2M}$) are compared to that of the WT in the presence of 2.0 M Gdn·HCl to yield the difference ($\Delta\Delta G^\circ_{u,2M}$). The estimated error in values of stability changes of mutants, relative to that of the WT protein ($\Delta\Delta G^\circ_{u,2M}$) is ±0.4 kcal/mol. Stabilities of mutants at positions 35, 47, 28, 64, 65, and 68 have been reported (4, 31, 38) and are taken from those works.

RESULTS AND DISCUSSION

Mutants of the Gene V Protein. The data compiled in Table 1 include stability and DNA binding affinity measurements for a variety of single mutants of the gene V protein and for a series of double mutants constructed by combination of

these single mutations. These data were used to explore the utility of concomitantly engineering *in vitro* stability and DNA binding affinity of the gene *V* protein through the use of additive mutational effects. Logically, this analysis consists of two steps: (*i*) assessment of the additivity of mutational effects on DNA binding and stability for the mutants listed in Table 1 and (*ii*) comparisons of the combined properties of a group of double mutants with those found in the starting group of single mutants.

Additivity of Mutational Effects on the Gene V Protein. It has been observed that the effects of accumulating mutations on *in vitro* properties of proteins are very nearly additive, provided that the substituted residues do not interact, either functionally (as in the case of catalytic residues) or by direct contact (25). When the substituted residues in multiply substituted proteins do interact or disrupt an interaction present in the WT protein, the effects are not additive. This is a potential stumbling block to protein engineering by accumulating mutations because the *in vitro* properties of such multiple mutants could not be predicted from the effects of the single mutants taken alone. Fortunately, interaction between residues in proteins appears to be relatively rare, with the exception of catalytic residues (25). We assessed the additivity of stability changes among single and double mutants of the gene *V* protein by considering the stabilities of double mutants, along with the stabilities of their constituent single mutants, and comparing these to the stability of the WT. Fig. 1 compares the stability changes, ($\Delta\Delta G^{\circ}_{u,2M}$) relative to WT, of the double mutants listed in Table 1 with the sums of the stability changes of their constituent single mutants. The data generally fall along a straight line with a slope near unity, demonstrating that interactions between substituting residues appear to be minimal in the combinations tested. This simple additivity was unexpected for pairwise substitutions of residues at positions 35 and 47 of the gene *V* protein due to the close proximity of these sites in a published crystallographic model of the protein (39). However, NMR results (40) and a revised crystal structure determined in our laboratory (Skinner *et al.*, unpublished data) show that, instead, Ile-47 is near Cys-33. This in turn may explain the significantly nonadditive stability effect of combining the mutation C33M with I47C (Table 1 and Fig. 1). This double mutant lies far from the line described by the rest of the combinations shown in Fig. 1.

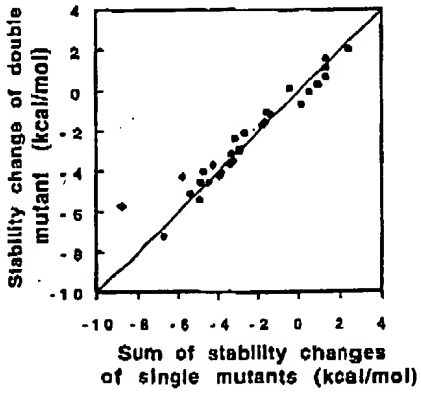

FIG. 1. Additivity of mutational effects on gene *V* protein stability. The stability change ($\Delta\Delta G^{\circ}_{u,2M}$), relative to the WT protein, of gene *V* protein double mutants is shown on the *y* axis. The *x* axis shows the sum of the stability changes, also relative to the WT protein, of the constituent single mutants. Positive values of $\Delta\Delta G^{\circ}_{u,2M}$ indicate proteins with increased stability. The combination of the mutants C33M and I47C is indicated by the diamond (♦). A line with unit slope is shown for reference.

Analogous to the results for stability changes, the DNA binding affinity changes ($\Delta\Delta G^{\circ}_{d,0.15M}$), relative to WT, of the double mutants tested are generally the sums of the binding affinity changes of their constituent single mutants, as shown in Fig. 2.

Presumably, some combinations of single mutations in addition to the C33M/I47C double mutant will lead to interactions between substituting residues. However, the frequent observation of simple additivity for the pairs of sites studied in gene *V* protein (Figs. 1 and 2), and the stepwise accumulation of stability changes observed in other proteins, including variants containing up to six substitutions (16, 25), suggests that interactions between substituents may be relatively rare as long as the sites chosen as targets for substitution are not obviously related by proximity or function (as in the case of catalytic residues). Thus, the additivity of effects of substitutions shown in Figs. 1 and 2 suggests that it should be possible to alter both gene *V* protein DNA binding affinity and stability in an additive and predictable fashion by combining previously characterized single mutants.

Engineering the Gene V Protein. To simultaneously adjust gene *V* protein DNA binding affinity and stability by multiple mutagenesis, it is important to know the relationship between the binding affinity changes and stability changes of the starting single mutants. If the two properties are strongly correlated, then the stability change caused by a mutation will always be in the same direction, relative to the WT, as the DNA binding affinity change, restricting the range available in one parameter (binding affinity or stability) relative to the other. On the other hand, if DNA binding affinity changes are loosely correlated or uncorrelated with stability changes in the starting group of single mutants, then it should be possible to generate mutants whose stability changes range widely with respect to DNA binding affinity changes. In this case, the DNA binding affinity and stability of the gene *V* protein can be altered simultaneously yet independently of each other, simply by combining single mutants to give the desired changes in each parameter.

To assess the correlation between DNA binding affinity changes and stability changes in the starting group of single mutants, $\Delta\Delta G^{\circ}_{d,0.15M}$ (the DNA binding affinity change) is plotted against $\Delta\Delta G^{\circ}_{u,2M}$ (the stability change) in Fig. 3*A*. Positive values of $\Delta\Delta G^{\circ}_{u,2M}$ or $\Delta\Delta G^{\circ}_{d,0.15M}$ indicate higher stability or increased DNA binding affinity, respectively.

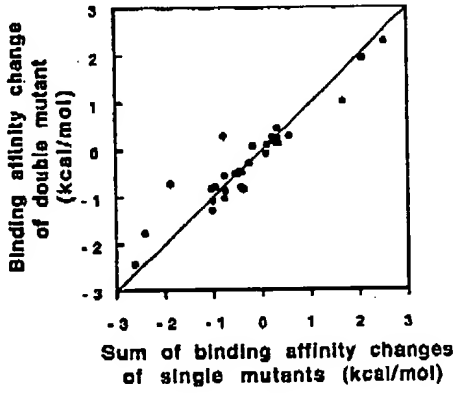

FIG. 2. Additivity of mutational effects on gene *V* protein DNA binding affinity. Binding affinity change ($\Delta\Delta G^{\circ}_{d,0.15M}$), relative to the WT protein, of gene *V* protein double mutants is shown on the *y* axis. The *x* axis shows the sum of the binding affinity changes, also relative to the WT protein, of the constituent single mutants. A positive value of $\Delta\Delta G^{\circ}_{d,0.15M}$ indicates enhanced binding to ssDNA relative to WT. The combination of the mutants C33M and I47C is indicated by the diamond (♦). A line with unit slope is shown for reference.

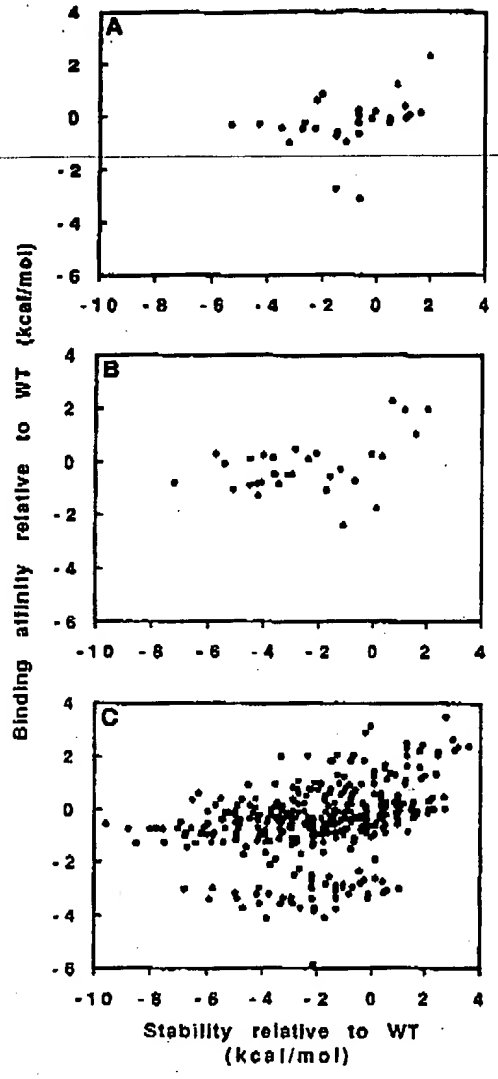

FIG. 3. (*A*) Comparison of ssDNA binding affinity changes ($\Delta\Delta G^{\circ}_{a,0.15M}$) with stability changes ($\Delta\Delta G^{\circ}_{u,2M}$) for single-substitution mutants shown in Table 1. The stability change, relative to the WT, is plotted on the *x* axis while DNA binding affinity change is plotted on the *y* axis. The upper right portion of the graph, therefore, contains mutants with higher stability and DNA binding affinity than the WT. (*B*) Comparison of ssDNA binding affinity changes with stability changes for double-substitution mutants shown in Table 1. (*C*) Calculated stability changes and ssDNA binding affinity changes resulting from all possible pairwise combinations of single mutants shown in Table 1, with assumptions as described in text.

than for the WT protein. Stability changes, relative to the WT protein, are only weakly correlated with binding affinity changes (also relative to WT), indicating that large changes of each parameter with respect to the other can be achieved by combining these mutants. Fig. 3*A* also shows graphically that very few of the single mutants alter stability without a concomitant DNA binding affinity change, or vice versa. Most of the single mutations decrease both stability and DNA binding affinity, but some mutations cause increases in one or both parameters relative to WT, indicating that adjustments of each parameter in either direction are potentially possible.

The results of combining single substitutions to alter two properties of the gene *V* protein are shown in Fig. 3*B*. The double mutants in Fig. 3*B* differ, as a group, in their properties from the starting single mutants in Fig. 3*A*, demonstrating that noncorrelated but additive changes in gene *V* protein DNA binding affinity and stability can be used to create a group of double mutants with distinctive *in vitro* properties. The double mutants shown in Fig. 3*B* are only a small fraction of the possible pairwise combinations of the single mutants shown in Table 1. Fig. 3*C* shows the predicted result, assuming that DNA binding affinity changes and stability changes are always directly additive (meaning no interactions between sites) of all possible pairwise combinations of the single mutants shown in Table 1. Simple pairwise combination of a starting group of 29 single-substitution mutants at 17 sites leads potentially to pairs of proteins differing by as much as 12 kcal/mol in stability or 8 kcal/mol in DNA binding affinity without substantial changes in the other parameter. Stability increases or DNA binding affinity increases relative to the WT of 3 kcal/mol or both are apparently within reach through pairwise combination of single mutants as well. Potential examples are the double mutants L32Y/L28V (increasing stability), I6V/F73W (increasing DNA binding affinity), and L32Y/F73W or F73W/M77I (increasing both parameters) (Table 1). Large destabilizations and reductions in DNA binding could also be achieved in theory. In practice, proteins with $\Delta\Delta G^{\circ}_{u,2M}$ more negative than ≈7.5 kcal/mol are substantially unfolded at 25°C and are difficult to produce *in vivo* (unpublished observations). Similarly, rapid purification of gene *V* protein variants employs ssDNA affinity chromatography, which may impose a lower limit on the obtainable reduction in DNA binding affinity. Within the practical limits of expression and purification, potentially hundreds of proteins with precisely engineered DNA binding affinities and stabilities could be produced by pairwise combination of a relative handful of single-substitution mutants.

Potential Utility of Protein Engineering by Multiple Mutation. Starting from a modest group of characterized single-substitution mutants, we have created double mutants with distinctive pairings of *in vitro* stability and DNA binding affinity. The stabilities and DNA binding affinities of these double mutants are generally additively related to the stability and DNA binding affinity changes of the starting single mutants. These results suggest that large numbers of proteins with precisely tailored properties can be deliberately constructed by the appropriate combination of single-substitution mutants. The properties of the double-substitution proteins can be predicted, based on those of the starting mutants, if care is taken to ensure (as much as possible) that the sites chosen for substitution will lead to simple additive effects (25). Simple additivity may not occur if the substituting residues contact each other, due to a change in the energy of interaction between the two sites (25). However, the potential interactions between amino acid residues, with the exception of charge–charge interactions, are strongly distance-dependent (41). Also, the effects of amino acid substitutions on protein structure are often localized to the immediate vicinity of the substitution (10, 42–45). It has been observed that the effects of substitutions on the properties of proteins are generally additive when the sites of substitution are not in van der Waals contact with each other (25). Consistent with these suggestions, nonadditivity of stability effects in the gene *V* protein is observed for mutations at sites that are close to each other (sites 33 and 47) but not for more distant sites (sites 35 and 47).

The ability to alter multiple properties of the gene *V* protein by combining substitutions is potentially useful in the further characterization of the gene *V* protein as well. For example, some pairwise combinations of substitutions could lead to proteins that differ in sequence and in some properties, yet that possess both WT stability and DNA binding affinity. Combinations of L32Y with V35I, I47L, or L65P might lead to proteins of this type (Table 1). These proteins could be

used to study the effects of substitutions on other properties of the gene *V* protein such as resistance to irreversible thermal denaturation (31) or folding and unfolding rates (30) in the context of WT stability and DNA binding affinity.

Protein engineering through the combination of single-substitution mutants may be most successful at adjusting those properties of concern *in vitro*, rather than *in vivo* activity. This is because activity *in vivo* may involve sequential interactions or parameters such as lifetime and folding/unfolding rates not considered in the *in vitro* analysis of the effects of substitutions. Mutations leading to DNA binding affinity and stability changes probably alter these other properties as well, complicating the task of engineering a particular property as the number of parameters to be maintained near WT values increases. Nevertheless, engineered proteins may find many applications *in vitro* where defects in some properties may be acceptable, and the ability to rapidly adjust the *in vitro* properties of proteins by combining well-characterized single substitutions should facilitate future protein engineering efforts.

1. Kellis, J. T., Jr., Nyberg, K., Sali, D. & Fersht, A. R. (1988) *Nature (London)* 333, 784–786.
2. Kellis, J. T., Jr., Nyberg, K. & Fersht, A. R. (1989) *Biochemistry* 28, 4914–4922.
3. Sandberg, W. S. & Terwilliger, T. C. (1989) *Science* 245, 54–57.
4. Sandberg, W. S. & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. USA* 88, 1706–1710.
5. Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1986) *Proteins* 1, 43–46.
6. Matsumura, M., Yasumura, S. & Aiba, S. (1986) *Nature (London)* 323, 356–358.
7. Matsumura, M., Becktel, W. J. & Matthews, B. W. (1988) *Nature (London)* 334, 406–410.
8. Wetzel, R., Perry, L. J., Baase, W. A. & Becktel, W. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 401–405.
9. Matsumura, M., Signor, G. & Matthews, B. W. (1989) *Nature (London)* 342, 291–293.
10. Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8237–8241.
11. Dao-Pin, S., Baase, W. A. & Matthews, B. W. (1990) *Proteins* 7, 198–204.
12. Zhang, X.-J., Baase, W. A. & Matthews, B. W. (1991) *Biochemistry* 30, 2012–2017.
13. Stearman, R. S., Frankel, A. D., Freire, E., Liu, B. & Pabo, C. O. (1988) *Biochemistry* 27, 7571–7574.
14. Shortle, D., Stites, W. E. & Meeker, A. K. (1990) *Biochemistry* 29, 8033–8041.
15. Bryan, P. N., Rollence, M. L., Pantoliano, M. W., Wood, J., Finzel, B. C., Gilliland, G., Howard, A. J. & Poulos, T. L. (1986) *Proteins* 1, 326–334.
16. Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L. & Bryan, P. N. (1989) *Biochemistry* 28, 7205–7213.
17. Russell, A. J. & Fersht, A. R. (1987) *Nature (London)* 328, 496–500.
18. Carter, P. & Wells, J. A. (1988) *Nature (London)* 332, 564–568.
19. Carter, P., Nilsson, B., Burnier, J. P., Burdick, D. & Wells, J. A. (1989) *Proteins* 6, 240–248.
20. Nelson, H. C. M. & Sauer, R. T. (1985) *Cell* 42, 549–558.
21. Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. & Estell, D. A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1219–1223.
22. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5167–5171.
23. Carter, P. & Wells, J. A. (1987) *Science* 237, 394–399.
24. Liao, H., McKenzie, T. & Hageman, R. (1985) *Proc. Natl. Acad. Sci. USA* 83, 576–580.
25. Wells, J. A. (1990) *Biochemistry* 29, 8509–8517.
26. Imanaka, T., Shibazaki, M. & Tagaki, M. (1986) *Nature (London)* 324, 695–697.
27. Scrutton, N. S., Berry, A. & Perham, R. N. (1990) *Nature (London)* 343, 38–43.
28. Terwilliger, T. C. (1988) *Gene* 69, 317–324.
29. Terwilliger, T. C. (1988) *Gene* 71, 41–47.
30. Liang, H. & Terwilliger, T. C. (1991) *Biochemistry* 30, 2772–2782.
31. Zabin, H. B. & Terwilliger, T. C. (1991) *J. Mol. Biol.* 219, 257–275.
32. Pretorius, H. T., Klein, M. & Day, L. A. (1975) *J. Biol. Chem.* 250, 9262–9269.
33. Veiko, N. N., Gromova, E. S. & Shabarova, Z. A. (1979) *Mol. Biol. (Engl. Transl.)* 13, 1136–1146.
34. Alma, N. C. M., Harmsen, B. J. M., de Jong, E. A. M., van der Ven, J. & Hilbers, C. W. (1983) *J. Mol. Biol.* 163, 47–62.
35. Bulsink, H., Harmsen, B. J. M. & Hilbers, C. W. (1985) *J. Biomol. Struct. Dyn.* 3, 227–247.
36. de Jong, E. A. M., Harmsen, B. J. M., Konings, R. N. H. & Hilbers, C. W. (1987) *Biochemistry* 26, 2039–2046.
37. Davis, N. G., Boeke, J. D. & Model, P. (1985) *J. Mol. Biol.* 181, 111–121.
38. Zabin, H. B. (1991) Ph.D. Thesis (The University of Chicago, Chicago).
39. Brayer, G. D. & McPherson, A. (1983) *J. Mol. Biol.* 169, 565–596.
40. Folkers, P. J. M., van Duynhoven, J. P. M., Jonker, A. J., Harmsen, B. J. M., Konings, R. N. H. & Hilbers, C. W. (1991) *Eur. J. Biochem.* 202, 349–360.
41. Fersht, A. (1985) *Enzyme Structure and Function* (Freeman, New York).
42. Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987) *Nature (London)* 330, 41–46.
43. Howell, E. E., Villafranca, J. E., Waren, M. S., Oatley, S. J. & Kraut, J. (1986) *Science* 231, 1123–1128.
44. Katz, B. A. & Kossiakoff, A. (1986) *J. Biol. Chem.* 261, 15480–15485.
45. Wilde, J. A., Bolton, P. H., Dell'Acqua, M., Hibler, D. W., Pourmottabbed, T. & Gerlt, J. A. (1988) *Biochemistry* 27, 4127–4132.

Selection for Tn*10* Tet Repressor Binding to *tet* Operator in *Escherichia coli*: Isolation of Temperature-Sensitive Mutants and Combinatorial Mutagenesis in the DNA Binding Motif

A. Wissmann,[1] L. V. Wray, Jr.,[2] U. Somaggio, R. Baumeister, M. Geissendörfer[3] and W. Hillen

Lehrstuhl für Mikrobiologie, Friedrich Alexander Universität Erlangen/Nürnberg, D-8520 Erlangen, Federal Republic of Germany

Manuscript received July 13, 1990
Accepted for publication February 8, 1991

ABSTRACT

We have constructed a genetic assay which selects positively for a functional interaction between Tet repressor and its cognate operator in *Escherichia coli*. In this strain Tet repressor blocks expression of *lacI* and *lacZ*. This leads to derepression of a *lacPO* controlled *galK* gene. The strain can be selected by growth on galactose as the sole carbon source and screened for the β-galactosidase phenotype. These features allow the identification of one candidate among 10^8 false clones on a single plate. The assay was applied to select mutants with a ts DNA binding phenotype and to screen oligonucleotide generated Tet repressor mutants. Analysis of these mutations revealed that they affect DNA and inducer binding and possibly the dimerization domains. These mutations are located at residues 21, 48, 49, 89 and at the C terminus of the protein (193), respectively.

THE *tet* determinant on transposon Tn*10* confers high level resistance to tetracycline in *Escherichia coli* and other enteric bacteria (FOSTER, HOWE and RICHMOND 1975; KLECKNER *et al.* 1975). Expression of resistance is regulated very tightly at the level of transcription (BECK *et al.* 1982). The regulatory region contains the *tet* promoters as well as two *tet* operators O_1 and O_2 (BERTRAND *et al.* 1983) which are bound by Tet repressor preventing transcription (HILLEN *et al.* 1983; WRAY and REZNIKOFF 1983; MEIER, WRAY and HILLEN 1988). The inducer tetracycline binds to Tet repressor leading to the loss of DNA binding activity. A special feature of this system is the opposite orientation of the *tetR* gene encoding Tet repressor relative to the resistance gene. *tetR* is transcribed by promoters within the *tet* regulatory region and is subject to autoregulation (BERTRAND *et al.* 1983; HILLEN, SCHOLLMEIER and GATZ 1984). These features are summarized in Figure 1.

Three essential Tet repressor functions are depicted in the figure: dimerization to form the active DNA binding form, DNA recognition and induction by tetracycline. *tet* operator binding probably makes use of an α-helix-turn-α-helix supersecondary structure (ISACKSON and BERTRAND 1985) and several mutants lacking inducibility by tetracycline have been mapped between amino acids 64 and 107 of the 207 amino acid primary structure (POSTLE, NGUYEN and BERTRAND 1984; SMITH and BERTRAND 1988). We are interested in studying the functional basis of Tet repressor activities and describe in this article the construction, efficiency and application of an *E. coli* strain that allows positive selection for functional Tet repressor-*tet* operator binding. Similar approaches have been used to analyze other protein-DNA recognition reactions (for example see ELLEDGE *et al.* 1989).

MATERIALS AND METHODS

Bacteria and phage: All bacterial strains are derived from *E. coli* K12. Strain R1291 (*pro galK2 rpsL srl::Tn10*) is a derivative of *E. coli* N99 and was obtained from B. RAK, Freiburg, Federal Republic of Germany. This strain was transduced to a *pro*$^+$Δ*lacX74* genotype using a P1 lysate grown on *E. coli* X7029 (BECKWITH and SIGNER 1966). The resulting *E. coli* strain WH205 was then transduced to *srl*$^+$ by a phage T4GT7 lysate (WILSON *et al.* 1979) derived from *E. coli* N100 (MCKENNEY *et al.* 1981). This yielded strains with a *srl*$^-$ *Tc*^{r} phenotype. Since *recA* can be cotransduced with *srl* by T4GT7, candidates were analyzed for hypersensitivity to UV. Isogenic strains WH206 (*galK2* Δ*lacX74 rpsL*) and WH207 (as WH206 but *recA*) were obtained which differ phenotypically only in their UV sensitivity. *E. coli* JM101 was used for propagation of M13mp9 phages and derivatives thereof (YANISCH-PERRON, VIEIRA and MESSING 1985). Phages λplac5 (IPPEN, SHAPIRO and BECKWITH 1971) and derivatives were propagated in *E. coli* XA105 (MILLER *et al.* 1977). *E. coli* NK5031(λtet50) was obtained from L. SMITH, San Diego. The lysogenic phage λtet50 is identical to λRStet158–50 (SMITH and BERTRAND 1988; L. D. SMITH, personal communication) and carries the wild-type *cI* allele as well as a *tetA-lacZ* fusion. Strain KD1067 (DEONEN and COX 1974) was used as a mutator strain for plasmid DNA. Plasmids were constructed and transformed to *E. coli* strains RRIΔZM15 (RÜTHER 1982) or X7029 (BECKWITH and SIGNER 1966).

[1] Present address: BASF Bioresearch Corporation, 195 Albany Street, Cambridge Massachusetts.
[2] Present address: Department of Microbiology, Boston University, 80 East Concord Street, Boston Massachusetts 02118-2394.
[3] Present address: Chemotherapeutisches Forschungsinstitut Georg Speyer Haus, Paul Ehrlich Strasse 42–44, 6000 Frankfurt/M.70, Federal Republic of Germany.

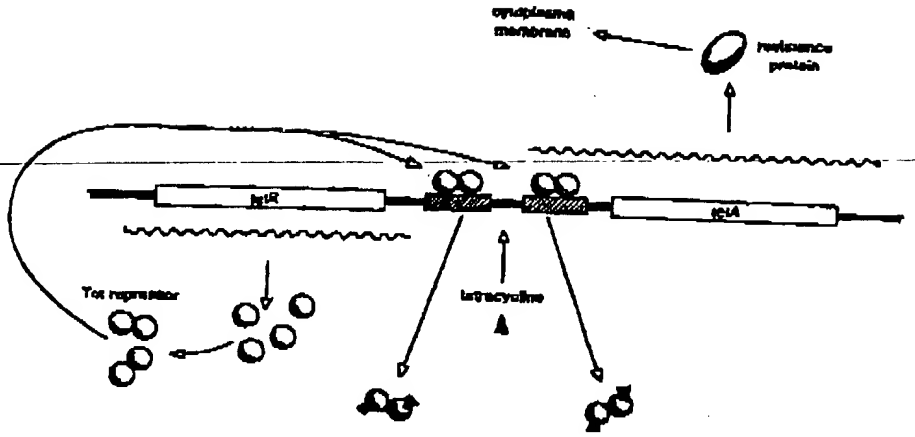

FIGURE 1.—Regulation of gene expression of the transposon Tn*10*-encoded tetracycline-resistance determinant. Both genes *tetA* (encoding the resistance protein) and *tetR* (encoding the Tet repressor) are indicated. Their divergent expression is symbolized by wavy lines corresponding to the respective mRNAs. The central *tet* regulatory region consists of several promoters (not shown) and the two *tet* operators O_1 and O_2 represented by hatched boxes. Tetracycline is indicated by the small rectangle which binds to and induces Tet repressor. The Figure was adapted from WISSMANN and HILLEN (1989).

DNA sequence analyses: Single-stranded M13 DNA and double-stranded plasmid DNA were sequenced according to the method of SANGER, NICKLEN and COULSON (1977) and HATTORI and SAKAKI (1986).

Determination of β-galactosidase activities: Assays were done exactly as described by MILLER (1972), except that cultures were grown in LB supplemented with the appropriate antibiotics. For induction studies overnight cultures were grown in the presence of 0.1 μg/ml tetracycline, whereas 0.2 μg/ml were added to log cultures. All measurements were repeated at least twice.

Media, enzymes and chemicals: Media and general phage techniques have been described (MILLER 1972; MANIATIS, FRITSCH and SAMBROOK 1982). Antibiotics and *o*-nitrophenyl-β-D-galactoside were obtained from Sigma, St. Louis. Restriction endonucleases, *E. coli* DNA polymerase I large fragment, T7 polymerase, calf intestine alkaline phosphatase and T4 DNA ligase were purchased either from New England Biolabs (Schwalbach), Pharmacia (Freiburg), Boehringer (Mannheim) or BRL (Dreieich). ATP, deoxyribonucleoside triphosphates and dideoxyribonucleoside triphosphates were obtained from Boehringer (Mannheim). [α-^{32}P]dATP (400 Ci/mmol) was purchased from Amersham (Braunschweig). Oligonucleotides were synthesized using an Applied Biosystems automated DNA synthesizer model 381A.

Molecular techniques: Mutagenesis of Tet repressor positions 46 to 49 was accomplished by mutually primed synthesis of degenerate oligonucleotides as detailed by HILL (1989). The sequence of the oligonucleotide was 5' GCCAGCATGTAAAAAATAAGCGGGCCCTGCTCG-ACGCGTCGAGC 3'. Bold letters (bases shown are wild type) indicate that 6–7% each of the three non-wild-type bases were added at these positions during synthesis of the oligonucleotide.

Plasmids: Plasmid pWH410 contains a fusion of the *tet* regulatory region to the *lac* operon (*tetA-lacZ* fusion). It was derived from pMC1403 (CASADABAN, CHOU and COHEN 1980) and allows Δlac *E. coli* strains to grow on lactose as the sole carbon source. Plasmid pWH414 differs in two aspects from pWH410. First, it carries a *tetR-lacI* fusion (Figure 2). Second, it contains a one base pair frameshift mutation at the fusion of *tetA* and *lacZ*. This renders Δlac *E. coli* strains unable to grow on lactose. Nevertheless, phenotypical detection of β-galactosidase activity with X-Gal is still possible.

Transdominance was analyzed in strains containing pWH855. This plasmid is a pBR322 derivative in which the *tet* regulatory region was deleted yielding pWH806 and the promoterless Tn*10 tetR* gene was inserted resulting in low level constitutive expression (MÜLLER-HILL, CRAPO and GILBERT 1968).

Plasmid pWH1411 was used for the cassette mutagenesis and as a derivative of pACYC177 (CHANG and COHEN 1978) is compatible to plasmids derived from pBR322. It confers resistance to chloramphenicol and contains a constitutively expressed *tetR* gene. To allow cloning of short oligonucleotide cassettes between singular restriction sites, the sequence of the *tetR* gene was altered without changing the encoded protein sequence. pRT240 is similar to pWH1411, except that it confers resistance to kanamycin and contains a wild-type *tetR* gene (BERTRAND *et al.* 1984; MEIER, WRAY and HILLEN 1988).

The pACYC177 derivatives pWH1200 and pWH1201 (ALTSCHMIED *et al.* 1988), pUC19 (YANISCH-PERRON, VIEIRA and MESSING 1985), pWH483 (MEIER, WRAY and HILLEN 1988) and pMc5–8 (STANSSENS *et al.* 1989) have been described. Plasmid pWH1012 (STEMORE *et al.* 1990) with divergent *tetR-galK* and *tetA-lacZ* transcriptional fusions was used for quantitative analyses of Tet repressor binding to *tet* operator *in vivo*.

Phage constructions and crosses: pWH483 was digested with *Nde*I and *Sma*I yielding a 1950-bp fragment with the entire *galK* gene. In addition, this fragment contains a segment of 180 bp with translational stops in all three reading frames 5' of the gene and a λt_0 terminator following the 3'-end of *galK*. After filling in the protruding ends the fragment was cloned into *Hinc*II linearized M13mp9. A candidate with *lac* dependent transcription of *galK* was named mWH22. A second *lac* operator with the proposed ideal binding sequence for Lac repressor (SADLER, SASMOR and BETZ 1983) was cloned 19 bp upstream of the start codon for *galK* into the single *Nru*I site of mWH22 yielding mWH25. In this construction palindromic centers of the two lac operators are separated by 233 bp. The *galK* construct from mWH25 was recombined into the *lac* sequences present on λplac5 to yield λWH25 (YU and REZNIKOFF 1984). Since this phage carries the cI^{857} allele from λplac5, *E. coli* strains lysogenized with this phage were grown at temperatures below 33°.

The construction of phage λtet50 has been described (SMITH and BERTRAND 1988). *E. coli* NK5031(λtet50) was treated with mitomycin C and the resulting phage lysate used to lysogenize *E. coli* WH207.

Selection of temperature-sensitive Tet repressor mutants: Mutagenized pRT240 was transformed to *E. coli*

WH207 containing pWH410 and grown to saturation at 42° in minimal medium with lactose. Since pWH410 contains a fusion of the *tet* regulatory region to the *lac* operon this step represents a selection against binding of Tet repressor to *tet* operator. Afterward pRT240 derivatives were isolated and retransformed to WH207(λWH25) containing pWH414. Transformants were grown to saturation at 28° in minimal medium with galactose. Here, cells containing *tet* operator bound by Tet repressor are selected. The pRT240 derivatives were isolated and retransformed to WH207 with pWH410. Again cells were grown to saturation at 42° in minimal medium using lactose. The pRT240 derivatives were isolated, transformed to WH207 with pWH410 and plated on glucose minimal medium supplemented with ampicillin, kanamycin and X-Gal. Plates were incubated at 42° for 2 days and blue colonies transferred to fresh plates containing the identical medium. *lacZ* phenotypes were scored after incubation at 28° for 2 days. *tetR* genes were recloned as *Hin*dIII fragments in pUC19. From derivatives with *lac* promoter *tet* fusions, *Eco*RI/*Sph*I DNA fragments containing *tetR* were then inserted into the respective sites of pWH1200 and pWH1201. This yielded two sets of plasmids with pWH1200 derivatives directing a "high," and pWH1201 derivatives directing a "low" level constitutive expression of *tetR in vivo* (BERTRAND *et al.* 1984).

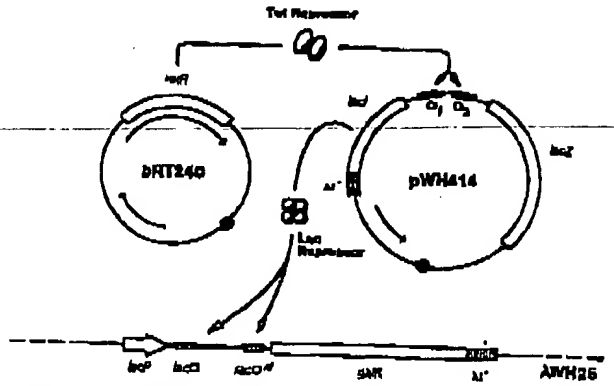

FIGURE 2.—Selection of Tet repressor binding to *tet* operator. DNA is indicated by thin lines, relevant genes as open boxes, the λ° terminators as stippled boxes, the *lac* operators as hatched boxes, the *tet* operators as filled boxes and the *lac* promoter on λWH26 as an open arrow. Filled circles mark the origins of replication of both plasmids. The arrow expanding through the *tetR* gene in pRT240 indicates the transcript originating from the *bla* promoter, whereas the other arrow defines the kanamycin resistance gene. The arrow in pWH414 indicates the *bla* gene. Tet repressor is shown as a dimer and Lac repressor as a tetramer.

RESULTS

Selection of Tet repressor binding to *tet* operator: The selection makes use of the *tet* directed expression of divergently arranged *lacZ* and *lacI* genes. Binding of repressor to the *tet* operators turns off transcription of both genes resulting in *lacZ*⁻ *E. coli* colonies. At the same time, the absence of Lac repressor allows expression of a galactokinase gene driven by the the *lac* regulatory region. This enables the *E. coli* strain to use galactose as the sole carbon source. In the absence of Tet repressor binding to *tet* operators, *lacZ* as well as *lacI* are expressed. Lac repressor binds to the *lac* operators and prevents transcription of *galK*. The cell cannot utilize galactose as the sole carbon source for growth and displays a *lacZ*⁻ phenotype.

The selection system consists of two plasmids and a λ prophage and is depicted in Figure 2. pWH414 makes use of the divergent *tet* regulatory region in that both a *tetR-lacI* transcriptional fusion as well as a *tetA-lacZ* fusion are present on the same plasmid. Tet repressor is supplied in *trans* by a second compatible plasmid (pRT240). The third component of the system is the prophage λWH25 which provides a single copy *lacPO-galK* fusion. The host strain is *E. coli* WH207 and has a *gal* operon with the *galK2* mutation (see MATERIALS AND METHODS).

A qualitative analysis of this system shows that all components behave as anticipated (see Table 1, lines 1 and 2). In the presence of Tet repressor, the strain is *gal*⁺ and *lacZ*⁻ (line 2, galactose alone). In the absence of Tet repressor, the strain is *gal*⁻ (line 1, galactose ± tetracycline; line 2, galactose + tetracycline). In the absence of Tet repressor, *lacI* repression can be partially alleviated by addition of isopropyl thiogalactoside (IPTG) (line 1, galactose + IPTG); complete derepression is probably not achieved because Lac repressor is present in such a high amount that it is never fully induced at the IPTG concentration used (10^{-3} M).

TABLE 1

Tet repressor and Lac repressor dependent expression of galactokinase in *E. coli* WH207(λWH25)

Plasmid	Tet repressor	Growth and phenotype of strains on: Glucose	Galactose	Galactose + IPTG	Galactose + tetracycline
pWH414	−	+b	−	+/−b	−
pWH414	+	+w	+w	+w	−
pWH414-2A	−	+b	−	+/−b	−
pWH414-2A	+	+b	−	+/−b	−

Strains of *E. coli* WH207(λWH25) with the indicated plasmids were streaked on minimal plates containing the indicated carbon source and inducer, as well as ampicillin, kanamycin and X-Gal. Plates were incubated at 30° for 3 days and scored for colony growth and color. Abbreviations used are "+" large single colonies, "+/−" small single colonies, "−" no single colonies but very thin bacterial film visible, "w" white colonies and "b" blue colonies. The presence of Tet repressor is indicated by "+" (pRT240), whereas the absence is indicated by "−" (pWH1200). The final concentrations of inducers were 10^{-3} M for IPTG and 0.5 μg/ml (corresponding to subinhibitory amounts) for tetracycline.

We have analyzed the selection system with an operator constitutive mutation to demonstrate the necessity of functional *tet* operators for the observed regulation. For this purpose pWH414–2A was used instead of pWH414 which differs from the latter by a total of 4-bp exchanges in the *tet* operators. MEIER, WRAY and HILLEN (1988) have shown that these mutations reduce binding of Tet repressor by about

three orders of magnitude. The phenotypes in the presence and absence of wild-type Tet repressor are as anticipated (see Table 1, lines 3 and 4). Growth on glucose yields *lacZ*$^+$ phenotypes while growth on galactose does not occur irrespective of the presence of Tet repressor. In the presence of galactose and IPTG this strain grows and is *lacZ*$^+$ (see above).

For a quantitative determination of the selection efficiency, mixtures of strains were grown on selective plates. These contained cells with the components shown in Figure 2, and an excess of cells in which either the repressor encoding plasmid pRT240 was replaced by the vector without *tetR*, or the wild type operators (pWH414) were replaced by their constitutive mutants (pWH414-2A). The results demonstrate that 30 cells with wild-type Tet repressor and *tet* operator can be efficiently selected on a single plate among 10^8 cells with either no Tet repressor or the *tet* operator mutation. No white colonies indicating repression of *lacZ* by Tet repressor are selected as false positives from 10^8 cells. The appearance of a few blue colonies might be due to spontaneous mutations of the *lacI* gene. It is the advantage of the divergent *tet* regulatory region that these candidates can be easily identified and discarded.

Temperature-sensitive Tet repressor mutants: Temperature-sensitive Tet repressor mutations were selected by their ability to confer growth on lactose at 42° and growth on galactose at 28° in appropriate *E. coli* strains (see MATERIALS AND METHODS). Seven parallel selections using individual preparations of pRT240 from the *E. coli* mutator strain KD1067 (DEGNEN and COX 1974) were carried through. Five of these selections yielded colonies which were blue at 42° and white at 28° with frequencies ranging from 2 to 85%. The *tetR* genes from one candidate of each of the seven selections were sequenced. The obtained mutations are displayed in Figure 3.

Temperature-sensitive Tet repressor mutants contained either a glycine to glutamic acid exchange at position 21 (GE21) or an an isoleucine to asparagine exchange at position 193 (IN193). The latter was independently selected four times. Another mutant (see Figure 3) isolated by a different approach contains an alanine to aspartic acid exchange at position 89 (AD89) and was included in the further *in vivo* analyses. The two mutants without a temperature sensitive phenotype were identical and had a C-terminal deletion (Δ141). The wild-type and mutant *tetR* genes were recloned resulting in two sets of plasmids directing either "high" or "low" level expression of *tetR*.

The mutants were assayed *in vivo* for repression of a *tetA-lacZ* fusion at 28°, 37° and 42°. Furthermore, inducibility by tetracycline and transdominante over wild type was tested. The results are presented in Table 2. Tet repressor mutants GE21, AD89 and IN193 display a clear temperature dependency of *lacZ*

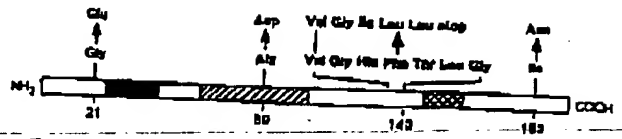

FIGURE 3.—Sequences of mutant Tet repressor proteins obtained from the selection for temperature sensitive variants. The Tet repressor protein with a total length of 207 amino acid residues is represented by a linear bar with both the N- and C-terminal ends indicated. The solid portion defines the potential α-helix-turn-α-helix motif, which is thought to be involved in DNA binding (amino acid residues 26 to 47; ISACKSON and BERTRAND 1985). The region of the protein for which mutants have been obtained that are defective for induction by tetracycline is hatched (amino acid residues 64 to 107; SMITH and BERTRAND 1988). Finally, a region of the protein that shows a high degree of variability when sequences of Tet repressor proteins from the five known resistance classes A through E are compared has been marked by crosshatching (amino acid residues 151 to 166; TOVAR, ERNST and HILLEN 1988). The glycine to glutamic acid exchange at position 21 is due to a transition of G to A, the exchange of isoleucine to asparagine at position 193 is the result of a T to A transversion and the deletion of one G in a run of four Gs leads to a frameshift resulting in a C-terminally deleted Tet repressor protein with a total length of 141 residues. Another temperature sensitive mutant which was isolated by a slightly different procedure (mutagenized pRT240 was transformed to *E. coli* X7029 containing plasmid pWH410 and resulting transformants analysed for their *lacZ* phenotype on X-Gal plates at 28° and 42°; M. GEISSENDÖRFER and W. HILLEN, unpublished results) was also included in the study. This mutant contains an exchange of alanine to aspartic acid at position 89 as the result of a C to A transversion.

repression, as evident from the ratios, whereas mutant Δ141 does not show repression in this system at all. At 28° and a "high" level of *tetR* expression IN193 shows almost wild-type activity and is clearly more active than AD89. On the contrary at a "low" level of *tetR* gene expression IN193 is not as effective as wild type and is even less active than AD89. The repression efficiencies encoded by the "high" expression plasmids are 95- and 900-fold higher for AD89 and IN193, respectively, than the ones found in the "low" expression plasmids. AD89 is only partially inducible by tetracycline, whereas the other mutants can be fully induced. GE21 and AD89 are transdominant.

Combinatorial mutagenesis at the C terminus of the putative DNA recognition α-helix of Tet repressor: Assuming that Tet repressor contains an α-helix-turn-α-helix motif for operator recognition (POSTLE, NGUYEN and BERTRAND 1984; PABO and SAUER 1984; ISACKSON and BERTRAND 1985), it is very likely that position 46 is part of the α-helix, whereas the secondary structures of residues 47 to 49 remain unclear. To gain information about their possible participation in operator binding a combinatorial cassette mutagenesis (REIDHAAR-OLSON and SAUER 1988) of Tet repressor was performed (see MATERIALS AND METHODS) as shown in Figure 4. Mutant plasmids were transformed to *E. coli* strains that either do or do not allow selection for *tet* operator binding of Tet repressor. *tetR* genes of candidates from both procedures were sequenced in the region of mutagenesis. Thirty-four

TABLE 2

In vivo analysis of mutant Tet repressors

Tet repressor	*tetR* expression	Repression 28°	37°	42°	Ratio 37°/28°	Tetracycline induction −Tc	+Tc	Transdominance −wt TetR	+wt TetR	Ratio +wt TetR/ 2.9
None		100.0 (±2.5)	100.0 (±3.4)	100.0 (±5.9)		100.0 (±5.1)	100.0 (±4.5)	100.0 (±4.1)	2.9 (±0.1)	1.0
Wild type	High	0.0 (±0.0)	0.1 (±0.1)	0.2 (±0.1)	1	1.3 (±0.2)	96.7 (±2.8)	1.1 (±0.0)	1.1 (±0.0)	0.4
GE21	High	18.4 (±2.1)	86.2 (±8.2)	99.3 (±1.2)	4.7	48.6 (±2.9)	104.8 (±4.7)	94.3 (±1.5)	18.5 (±1.5)	6.4
AD89	High	0.8 (±0.1)	45.2 (±3.1)	73.2 (±5.6)	54.0	4.6 (±1.3)	16.6 (±1.1)	73.4 (±0.2)	5.7 (±0.3)	2.0
IN193	High	0.1 (±0.1)	20.3 (±1.8)	69.2 (±1.5)	20.3	1.3 (±0.3)	98.7 (±2.4)	9.4 (±0.2)	1.6 (±0.0)	0.6
Δ141	High	99.4 (±1.9)	104.4 (±6.8)	100.0 (±0.4)	1.1			97.2 (±4.5)	2.5 (±0.2)	0.9
Wild type	Low	29.7 (±3.0)	51.7 (±1.9)	62.8 (±6.1)	1.7					
GE21	Low	96.0 (±2.7)	100.1 (±1.1)	100.8 (±5.9)	1.0					
AD89	Low	76.3 (±0.7)	88.3 (±6.9)	93.7 (±0.7)	1.2					
IN193	Low	89.4 (±1.1)	96.1 (±7.8)	97.4 (±4.1)	1.1					

β-Galactosidase determinations were performed in *E. coli* WH207 containing *tetA-lacZ* fusions and plasmids encoding the given Tet repressors. It is indicated whether the "high" or the "low" expression system for Tet repressor was used (see MATERIALS AND METHODS for details). β-Galactosidase values obtained in strains lacking Tet repressor were defined as 100%. "Repression" was measured using plasmid pWH1012 for the *tetA-lacZ* fusion. Actual values obtained were 275.8 (±8.7) units at 28°, 386.1 (±11.0) units at 37° and 341.7 (±15.7) units at 42°. Overnight cultures used for the inoculation of log cultures were also grown at the temperatures indicated, except for measurements at 42°, where overnight cultures were grown at 37°. When overnight cultures of strains containing mutants AD89 and IN193 in the "high" *in vivo* expression system were grown at 28° and log cultures were then grown at 42°, the percentages obtained for AD89 were 37.4% (±3.4%), whereas 2.7% (±0.1%) were obtained for IN193. "Tetracycline induction" was assayed using the prophage λtet50 for the *tetA-lacZ* fusion. Overnight and log cultures were grown at 28°. Actual values obtained in the absence of Tet repressor without addition of tetracycline were 4970 (±585) units and 4125 (±190) units in the presence of tetracycline. "Transdominance" was also determined using the prophage λtet50 as a *tetA-lacZ* fusion. In addition to the indicated plasmids, cells contained a second compatible plasmid which either was pWH806 (indicated by "−wt TetR") or pWH853 (indicated by "+wt TetR"). Details on both plasmids are given in MATERIALS AND METHODS. Overnight and log cultures were grown at 37°. Percentages are related to the 100% value defined by the strain lacking both wildtype and mutant Tet repressors. Typically, 3895 (±160) units were obtained in the absence of Tet repressor.

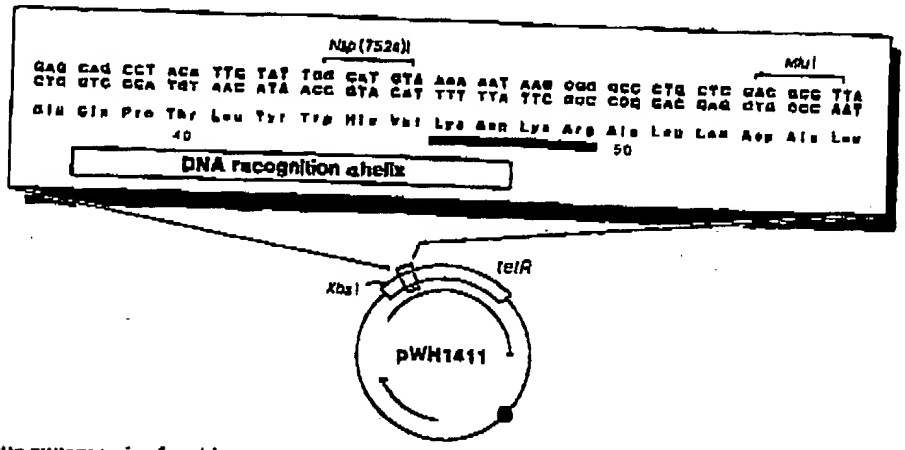

FIGURE 1.—Cassette mutagenesis of positions 46 to 49 of Tet repressor. The DNA of plasmid pWH1411 (see MATERIALS AND METHODS) is shown as a circle with the *tetR* gene emphasized by the open box. The origin of replication is depicted as a filled circle, the chloramphenicol acetyltransferase gene is indicated by the shorter arrow and the transcription originating from the *bla* promoter which leads to constitutive expression of *tetR* is marked by the longer arrow. The orientation of the *tetR* gene is indicated by the *Xba*I site (this restriction site is localized at the 5′ end of the gene). At the top of the figure, part of the DNA sequence of the *tetR* gene is shown together with the respective protein sequence. The localization of the potential DNA recognition α-helix is indicated by the open box below the sequence. Amino acid residues at positions 46 to 49 that were altered are underlined by a black bar. Singular restriction sites used for cloning are indicated above the DNA sequence.

different mutants with either single or multiple exchanges at positions 46 to 49 were obtained and analyzed *in vivo* for repression of a *tetA-lacZ* fusion at 28° and 37° and for tetracycline induction.

All mutants isolated with selection for Tet repressor binding to *tet* operator give rise to wild-type *lacZ* repression at 37°. The only exception was a triple mutant which showed a significant derepression of *lacZ*. At 28°, which was the temperature used for mutant selection, this candidate also displayed wild-type activity.

Single amino acid exchanges at positions 46 and 47 had no detectable effect on repressor activity (data not shown). Three of the five mutants at position 48

TABLE 3

Mutational analysis of Tet repressor positions 46 to 49

Tet repressor	Repression 28°	Repression 37°	Induction +Tetracycline
None	100.0 (±6.6)	100.0 (±5.6)	100.0 (±2.9)
Wild type	1.7 (±0.1)	1.4 (±0.2)	104.7 (±7.4)
KR48	ND	1.6 (±0.1)	ND
KQ48	5.0 (±0.4)	16.5 (±0.6)	105.7 (±6.2)
KH48	1.8 (±0.2)	5.6 (±1.0)	ND
KM48	ND	1.5 (±0.1)	ND
KT48	15.8 (±0.6)	45.2 (±5.5)	106.2 (±2.1)
RQ49	ND	1.5 (±0.1)	10.3 (±0.3)
RG49	ND	1.4 (±0.2)	ND
RP49	ND	1.3 (±0.1)	52.4 (±1.8)
RW49	ND	1.5 (±0.1)	103.6 (±1.1)

β-Galactosidase determinations were performed in *E. coli* WH207 (λtet50) with plasmids encoding the given Tet repressors. Values are given as percentages with regard to the amount of β-galactosidase measured for this strain containing plasmid pMc5-8 under the specific experimental conditions (specified as "none" in the table). Measurements were carried out with strains grown at 28° and 37°, with the respective overnight cultures grown at the same temperatures. Induction with tetracycline was also done at 37° (for details see MATERIALS AND METHODS).

showed a lower than wild-type repression activity (Table 3). The mutants at position 49 did not affect repression efficiencies but two candidates displayed only partial inducibility by tetracycline. Multiple amino acid exchanges at positions 46 to 49 influenced the repression activity only if position 48 was altered and the tetracycline inducibility only if position 49 was altered (data not shown).

DISCUSSION

Selection of Tet repressor binding to *tet* operator: The selection described above is very efficient, because single cells with wild-type Tet repressor binding to wild-type *tet* operator are found among a vast excess of up to 10^{11} cells with either no or reduced binding of Tet repressor to *tet* operator on one plate. The results with the 2A *tet* operator mutation show that Tet repressors must have an association constant of greater than $4 \times 10^{8}\ M^{-1}$ to *tet* operator in order to be selectable in this system.

Temperature-sensitive Tet repressor mutants: As depicted in Figure 3, GE21 is located in close proximity to the proposed α-helix-turn-α-helix element. It is the weakest DNA binder and shows the strongest transdominant phenotype of all the mutants analysed in this study. This mutant has been isolated previously by ISACKSON and BERTRAND (1985), but the authors did not describe the temperature dependent effect we have observed. We speculate that this mutation may interfere with the positioning of the DNA binding motif.

AD89 is located in a region where noninducible mutants have been mapped previously (SMITH and BERTRAND 1988). In agreement with these results it shows only partial induction by tetracycline but also a transdominant phenotype. At the same position SMITH and BERTRAND (1988) have also isolated a mutant (alanine to glycine) which allows only partial induction by tetracycline. Since the residue at position 89 affects both the DNA- and the inducer-binding domain it may be involved in structurally transmitting the signal of inducer binding to the DNA recognition domain.

Mutant IN193 is located in the C terminus, to which no function has been assigned so far. It gives rise to the strongest temperature-dependent effect observed in the course of this study. Tetracycline inducibility as far as detectable in our system is not affected and transdominance cannot be observed. When overnight cultures for β-galactosidase determinations were grown at 28° and log cultures were incubated at 42° mutant IN193 retains a much higher efficiency in *lacZ* repression than AD89 (see footnotes to Table 2). This phenotype corresponds to the *ts1* ("temperature-sensitive synthesis") mutants first described by SADLER and NOVICK (1965), where the oligomerized protein retains function upon shifting the culture to the nonpermissive temperature. Assembly of new dimers is inhibited at the nonpermissive temperature due to either a defect in folding of the monomer or inhibition of dimer formation (GOLDENBERG 1988). This might indicate that IN193 dimers already formed at 28° are not inactivated upon raising the temperature to 42°. On the contrary, it has been shown *in vitro* for mutant AD89 that shifting the temperature to 42° clearly inactivates the protein (B. STADE and W. HILLEN, unpublished results). Western blot analyses have shown identical levels of wild type and IN193 when grown at 28° while at 37° no IN193 protein is detectable (C. BERENS and W. HILLEN, manuscript to be published). Taken together with the large increase in repression with concentration (see Table 2) this leads us to speculate that position 193 of Tet repressor might be involved in dimer formation. The C termini of Tet repressor proteins from five resistance classes are rather homologous. They are preceded by a hypervariable region (amino acid residues 151 to 166 of Tn*10* Tet repressor; see Figure 3 and TOVAR, ERNST and HILLEN 1988) which could indicate a possible C-terminal dimerization domain of Tet repressor.

Tet repressor mutants at positions 46 to 49: Several of the Tet repressor mutants at position 48 show reduced DNA binding activity. This suggests that Lys^{48} either directly contacts DNA or that it participates in adjusting the structural conformation of the DNA recognition α-helix. Mutants at position 49 of Tet repressor show wild-type DNA binding, but in some inducibility with tetracycline is reduced. This phenotype can result from three effects: (i) reduced

binding of inducer; (ii) interference with the conformational change needed to transmit the signal of inducer binding to the DNA binding domain or (iii) increased affinity for operator resulting in a superrepressor. This effect is not found at the three other positions. The result is particularly surprising, since previously identified mutations in Tet repressor with a noninducible phenotype map between positions 64 and 107 (SMITH and BERTRAND 1988), in a region clearly distinct from the proposed DNA recognition α-helix. The large number of mutants, and a demonstration that some show reduced binding of tetracycline *in vitro* suggest that this region contains the binding site for tetracycline. Thus, Arg^{40} of Tet repressor might be located at the "DNA side" of the switch mediating inducer binding to the DNA binding site. However, superrepression as a result of additional interactions either stabilizing the repressor-operator or destabilizing repressor-nonoperator complexes (HECHT and SAUER 1985) is also possible. In conclusion, the combinatorial mutagenesis suggests that Lys^{48} may be involved in operator binding and Arg^{49} could be active in induction while no functions can be detected for Lys^{46} and Asn^{47}.

We thank B. RAK for plasmids, *E. coli* strains and his advice in strain construction. We also thank L. SMITH for *E. coli* NK5031(λtet50). We are grateful to G. HOFHAUS for help with some plasmid constructions and A. SKERWOTKA for excellent technical assistance. This work was supported by the Humboldt Siftung (L.V.W.), the Deutsche Forschungsgemeinschaft and the Fonds der chemischen Industrie.

LITERATURE CITED

ALTSCHMIED, L., R. BAUMEISTER, K. PFLEIDERER and W. HILLEN, 1988 A threonine to alanine exchange at position 40 of Tet repressor alters the recognition of the sixth base pair of tet operator from GC to AT. EMBO J. 7: 4011–4017.

BECK, C. F., R. MUTZEL, J. BARBÉ and W. MÜLLER, 1982 A multifunctional gene (*tetR*) controls Tn*10*-encoded tetracycline resistance. J. Bacteriol. 150: 633–642.

BECKWITH, J. R., and E. R. SIGNER, 1966 Transposition of the *lac* region of *Escherichia coli*. I. Inversion of the *lac* operon and transduction of *lac* by ϕ80. J. Mol. Biol. 19: 254–265.

BERTRAND, K. P., K. POSTLE, L. V. WRAY, JR., and W. S. REZNIKOFF, 1983 Overlapping divergent promoters control expression of Tn*10* tetracycline resistance. Gene 23: 149–156.

BERTRAND, K. P., K. POSTLE, L. V. WRAY, JR., and W. S. REZNIKOFF, 1984 Construction of a single-copy promoter vector and its use in analysis of regulation of the transposon Tn*10* tetracycline resistance determinant. J. Bacteriol. 158: 910–919.

CASADABAN, M. J., J. CHOU and S. N. COHEN, 1980 In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. 143: 971–980.

CHANG, A. C. Y., and S. N. COHEN, 1978 Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. 134: 1141–1156.

DEGNEN, G. E., and E. C. COX, 1974 Conditional mutator gene in *Escherichia coli*: isolation, mapping and effector studies. J. Bacteriol. 117: 477–487.

ELLEDGE, S. E., P. SUGIONO, L. GUARENTE and R. W. DAVIS, 1989 Genetic selection for genes encoding sequence-specific DNA-binding proteins. Proc. Natl. Acad. Sci. USA 86: 3689–3693.

FOSTER, T. J., T. C. B. HOWE and K. M. V. RICHMOND, 1975 Translocation of the tetracycline resistance determinant from R100-1 to the *Escherichia coli* K-12 chromosome. J. Bacteriol. 124: 1159–1158.

GOLDENBERG, D. P., 1988 Genetic studies of protein stability and mechanisms of folding. Annu. Rev. Biophys. Biophys. Chem. 17: 481–507.

HATTORI, M., and Y. SAKAKI, 1986 Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152: 232–238.

HECHT, M. H., and R. T. SAUER, 1985 Phage lambda repressor revertants. Amino acid substitutions that restore activity to mutant proteins. J. Mol. Biol. 186: 53–63.

HILL, D. E., 1989 Mutagenesis with degenerate oligonucleotides: creating numerous mutations in a small DNA sequence, pp. 8.2.1–8.2.7 in *Current Protocols in Molecular Biology*, Vol. 1, edited by F. M. AUSUBEL, R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J. A. SMITH and K. STRUHL. John Wiley & Sons, New York.

HILLEN, W., K. SCHOLLMEIER and C. GATZ, 1984 Control of expression of the Tn*10*-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the *tet* operon regulatory region. J. Mol. Biol. 172: 185–201.

HILLEN, W., C. GATZ, L. ALTSCHMIED, K. SCHOLLMEIER and I. MEIER, 1983 Control of expression of the Tn*10*-encoded tetracycline resistance genes. I. Equilibrium and kinetic investigations of the regulatory reactions. J. Mol. Biol. 169: 707–721.

IPTEN, K., J. A. SHAPIRO and J. R. BECKWITH, 1971 Transposition of the *lac* region to the *gal* region of the *Escherichia coli* chromosome: isolation of λ*lac* transducing bacteriophages. J. Bacteriol. 108: 5–9.

ISACKSON, P. J., and K. P. BERTRAND, 1985 Dominant negative mutations in the Tn*10 tet* repressor: evidence for use of the conserved helix-turn-helix motif in DNA binding. Proc. Natl. Acad. Sci. USA 82: 6226–6230.

KLECKNER, N., R. K. CHAN, B.-K. TYE and D. BOTSTEIN, 1975 Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J. Mol. Biol. 97: 561–575.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning—A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MCKENNEY, K., H. SHIMATAKE, D. COURT, U. SCHMEISSNER, C. BRADY and M. ROSENBERG, 1981 Gene amplification and analysis, p. 383–415 in *Structural Analysis of Nucleic Acids*, edited by J. G. CHIRIKJIAN and T. S. PAPAS. Elsevier/North-Holland, New York.

MEIER, I., L. V. WRAY and W. HILLEN, 1988 Differential regulation of the Tn*10*-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem *tet* operators O_1 and O_2. EMBO J. 7: 567–572.

MILLER, J. H., 1972 *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MILLER, J. H., D. GANEM, P. LU and A. SCHMITZ, 1977 Genetic studies of the Lac repressor. I. Correlation of mutational sites with specific amino acid residues: construction of a colinear gene-protein map. J. Mol. Biol. 109: 275–301.

MÜLLER-HILL, B., L. CRAPO and W. GILBERT, 1968 Mutants that make more *lac* repressor. Proc. Natl. Acad. Sci. USA 59: 1259–1264.

PABO, C. O., and R. T. SAUER, 1984 Protein-DNA recognition. Annu. Rev. Biochem. 53: 293–321.

POSTLE, K., T. T. NGUYEN and K. P. BERTRAND, 1984 Nucleotide

sequence of the repressor gene of the Tn*10* tetracycline resistance determinant. Nucleic Acids Res. 12: 4849–4863.

REIDHAAR-OLSON, J. F., and R. T. SAUER, 1988 Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241: 53–57.

RÜTHER, U., 1982 pUR250 allows rapid chemical sequencing of both DNA strands of its inserts. Nucleic Acids Res. 10: 5765–5772.

SADLER, J. R., and A. NOVICK, 1965 The properties of repressor and the kinetics of its action. J. Mol. Biol. 12: 305–327.

SADLER, J. R., H. SASMOR and J. L. BETZ, 1983 A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. Proc. Natl. Acad. Sci. USA 80: 6785–6789.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463–5467.

SIZEMORE, C., A. WISSMANN, U. GÜLLAND and W. HILLEN, 1990 Quantitative analysis of Tn*10* Tet repressor binding to a complete set of *tet* operator mutants. Nucleic Acids Res. 18: 2875–2880.

SMITH, L. D., and K. P. BERTRAND, 1988 Mutations in the Tn*10* tet repressor that interfere with induction; location of the tetracycline-binding domain. J. Mol. Biol. 203: 949–959.

STANSSENS, P., C. OPSOMER, Y. M. MCKEOWN, W. KRAMER, M. ZABEAU and H.-J. FRITZ, 1989 Efficient oligonucleotide-directed construction of mutations in expression vectors by the gapped duplex DNA method using alternating selectable markers. Nucleic Acids Res. 17: 4441–4454.

TOVAR, K., A. ERNST and W. HILLEN, 1988 Identification and nucleotide sequence of the class E *tet* regulatory elements and operator and inducer binding of the encoded purified *tet* repressor. Mol. Gen. Genet. 215: 76–80.

WILSON, G. G., K. K. Y. YOUNG, G. J. EDLIN and W. KÖNIGSBERG, 1979 High-frequency generalised transduction by bacteriophage T4. Nature 280: 80–82.

WISSMANN, A., and W. HILLEN, 1989 Tetracyclin-Resistenzdeterminanten: Mechanismen der Resistenz und Regulation ihrer Expression). Forum Mikrobiol. 12: 292–299.

WRAY, L. V., JR., and W. S. REZNIKOFF, 1983 Identification of repressor binding sites controlling expression of tetracycline resistance encoded by Tn*10*. J. Bacteriol. 156: 1188–1191.

YANISCH-PERRON, C., J. VIEIRA and J. MESSING, 1985 Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119.

YU, X. M., and W. S. REZNIKOFF, 1984 Deletion analysis of the CAP-cAMP binding site of the *E. coli* lactose promoter. Nucleic Acids Res. 12: 5449–5464.

Communicating editor: N. KLECKNER

Protein Engineering vol.6 no.3 pp.327–331, 1993

Recursive ensemble mutagenesis

Simon Delagrave, Ellen R.Goldman and Douglas C.Youvan[1]

Massachusetts Institute of Technology, Department of Chemistry, Room 56-213, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[1]To whom correspondence should be addressed

We have developed a generally applicable experimental procedure to find functional proteins that are many mutational steps from wild type. Optimization algorithms, which are typically used to search for solutions to certain combinatorial problems, have been adapted to the problem of searching the 'sequence space' of proteins. Many of the steps normally performed by a digital computer are embodied in this new molecular genetics technique, termed recursive ensemble mutagenesis (REM). REM uses information gained from previous iterations of combinatorial cassette mutagenesis (CCM) to search sequence space more efficiently. We have used REM to simultaneously mutate six amino acid residues in a model protein. As compared to conventional CCM, one iteration of REM yielded a 30-fold increase in the frequency of 'positive' mutants. Since a multiplicative factor of similar magnitude is expected for the mutagenesis of additional sets of six residues, performing REM on 18 sites is expected to yield an exponential (30 000-fold) increase in the throughput of positive mutants as compared to random $[NN(G,C)]_{18}$ mutagenesis.

Key words: light harvesting II/protein engineering/random mutagenesis

Introduction

Current endeavors to engineer new specificities in antibodies and their derivatives hold the promise of new therapeutic and diagnostic tools. The generation of new and informative mutant proteins is necessary to our understanding of protein structure–function relationships. Such tasks are made difficult by our inability to predict structure from primary sequence or even to predict function from structure. One strategy circumventing the gaps in our understanding involves the selection of desired phenotypes from a large pool of different genotypes, in a manner analogous to natural selection. A limitation of this is the combinatorial explosion problem: as the number of randomized (mutated with all 20 amino acids) sites in a protein increases, the number of possible combinations which must be evaluated to identify 'positives' grows exponentially as 20^n, where n is the number of sites altered. Ingenious methods have been devised to allow screening of increasingly complex libraries of mutant proteins, peptides and oligonucleotides. Phage display libraries (Smith, 1985; Hoogenboom *et al.*, 1991; Kang *et al.*, 1991) and mutated ribozyme populations (Beaudry and Joyce, 1992) are instances of 'systems' where the genotypes and phenotypes are physically linked to allow for rapid selection and amplification of extremely complex ensembles of mutants. To completely screen a library of mutant proteins with 20 randomized amino acid residues ($n = 20$), the synthesis of 20^{20} or 10^{26} different protein molecules is required. Obviously, this will challenge our technical capabilities for some time. It may be desirable to avoid the very high proportion of non-functional proteins in a random library and simply enhance the frequency of functional proteins, thus decreasing the complexity required to achieve a useful sampling of sequence space. Recursive ensemble mutagenesis (REM) is an algorithm which enhances the frequency of functional mutants in a library when an appropriate selection or screening method is employed (Arkin and Youvan, 1992a; Youvan *et al.*, 1992).

REM uses successive rounds of CCM (Oliphant *et al.*, 1986; Reidhaar-Olson *et al.*, 1991) to generate a diverse library of genetically altered proteins that fit certain selection criteria (Figure 1). Amino acids are retained in the library if they are found in an altered protein fitting the selection criteria. Lists of all amino acids that are acceptable at each mutated position (i.e. 'target sets' of amino acids) are compiled. In the next iteration of REM, combinatorial cassettes are resynthesized according to mathematical functions that bias the nucleotide mixtures (Arkin and Youvan, 1992b; Youvan *et al.*, 1992) at each mutated

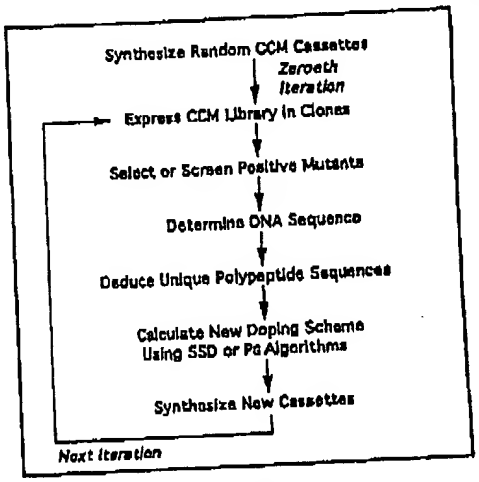

Fig. 1. REM involves the recursive use of combinatorial cassette mutagenesis (CCM). The first step of REM begins by expressing and screening a CCM library. Two or more 'positive' mutants are then picked and sequenced. (Positive mutants are defined in the current experiment as binding significant levels of red-shifted Bchl which is characteristic of LHII assembly.) Next, a list of unique protein sequences is determined by translating these DNA sequences. A 'unique sequence' is defined at the protein level. If more than one protein has the same sequence, only the first occurrence of this sequence is retained and counted as unique. For each mutated position in the protein, a target set of acceptable amino acids is compiled and the most appropriate dope is determined by a mathematical function such as group probability (P_α). The next iteration of REM proceeds by using these 'intelligent' dopes to generate a combinatorial cassette of lower complexity. In order to take advantage of the properties of REM, the complexity of the possible peptide sequences arising from CCM should be shown to be in vast excess of the screening size (Youvan *et al.*, 1992).

position in the protein to encode these target sets of amino acids. For example, if Ala, Ser and Thr occur at a given position in different selected mutants, these amino acids constitute the target set at that position. A mathematical function is used to select the best 'dope' that maximizes the probabilities of the amino acids in the target set. The next cassette is then designed such that this target set is encoded by a simple mixture of nucleotides at that codon [e.g. [(G,A,T)(C)(G,C)]]. In certain cases, where there is a good match between selection criteria and structure inherent in the genetic code (Sjostrom and Wold, 1985; Youvan, 1991) such as hydropathy and molar volume, computer simulations predict that multiple iterations of REM will yield thousands of times more mutants than conventional CCM (Arkin and Youvan, 1992b; Youvan *et al.*, 1992).

As a model system to experimentally verify the computer predicted amplification by REM, the light harvesting II (LHII) β-subunit gene (Youvan and Ismail, 1985) of *Rhodobacter capsulatus* was chosen. The LHII protein has two characteristic absorption bands in the near infrared (800 and 858 nm) that are red shifted relative to protein-free bacteriochlorophyll (Bchl) absorption at 770 nm. These prosthetic groups serve as colorimetric indicators of protein expression and subunit assembly. Six carboxy-terminal residues of the β-subunit were initially mutated by construction of a combinatorial cassette containing the sequence $[NN(G,C)]_6$, where 'N' designates an equiprobable mixture of all four nucleotides. This CCM library was conjugated into a strain of *R.capsulatus* (U71) totally deficient in Bchl-binding proteins or any other compounds with significant absorption in the near infrared (Youvan *et al.*, 1985). This deletion background facilitates the use of digital imaging spectroscopy (DIS) (Arkin *et al.*, 1990; Arkin and Youvan, 1993) to screen thousands of colonies directly on Petri dishes for LHII expression. We then sequenced five functional mutants and used this limited data to construct a new CCM library. The frequency of positives was increased 30-fold relative to the original library.

Materials and methods

Plasmids and strains

Plasmid pU4b is a shuttle vector used for cassette mutagenesis as well as expression of the mutant LHII genes (Goldman and Youvan, 1992). M13 was our vector for single-stranded sequencing and was propagated in *Escherichia coli* MV1190. *Escherichia coli* strain S17-1 was used for library construction and conjugation with *R.capsulatus* U71. For expression of the libraries, *R.capsulatus* U71, an LHII chromosomal deletion background (LHII and reaction center expression inactivated by a point mutation) was used.

Materials and DNA manipulations

DNA manipulations were essentially performed as described by Sambrook *et al.* (1989). Restriction enzymes were obtained from New England Biolabs, T4 DNA ligase was from Bethesda Research Labs as was Taq polymerase. Sequencing was carried out using a Sequenase kit from United States Biochemicals. Electroporation was carried out in 0.2 cm cuvettes on 0.45 ml of competent cells using a Bio-Rad electroporator according to instructions provided. All oligonucleotides were synthesized on an Applied Biosystems model 381 DNA synthesizer using commercially available reagents.

Library construction

The unique *Kpn*I and *Xho*I sites of pU4b flank the region encoding the dimer Bchl binding site and the carboxy-terminus of the β-subunit LHII gene. These restriction sites were engineered to allow double-stranded combinatorial cassettes to be subcloned in place of the wild type sequence.

Table 1. Sequences and corresponding phenotypes of mutants isolated from the zero and first iterations of REM

Mutant	Deduced sequence	OD 800 nm	OD 855 nm	OD855/OD800
Wild type	ATPWLG	0.26	0.39	1.5
REM0.6	LTPWVA	0.15	0.24	1.6
REM0.7	LTPWVP	0.13	0.20	1.5
REM0.8	ASPWMS	0.09	0.15	1.7
REM0.9	SSPWLP	0.15	0.22	1.5
REM0.10	FVWPGF	0.05	0.09*	1.8
REM1.1	STPWVF	0.11	0.17	1.5
REM1.2	FTPWVG	0.11	0.18	1.6
REM1.3	ATPWLA	0.10	0.15	1.5
REM1.4	STPWLA	0.33	0.48	1.5
REM1.5	LTPWGR	0.09	0.13	1.4
REM1.6	VTPWLP	0.11	0.18	1.6
REM1.7	VTPWLG	0.13	0.21	1.8
REM1.8	LIWPVL	0.05	0.09*	1.8
REM1.9	ALWPLV	0.05	0.09*	1.8
REM1.10	LTPWGG	0.20	0.29	1.5
REM1.11	VTPWVR	0.06	0.11	1.8
REM1.12	VTPWGL	0.12	0.21	1.8

*Peak shifted to 845 nm.

The sense strand of the 113-mers, which included the *Kpn*I–*Xho*I sites, as well as two PCR primers (20-mers each spanning a restriction site) were synthesized. The doped sequence within the cassette used in the zero iteration was $[NN(G,C)]_6$. The purified 113-mer was amplified by PCR. Amplified double-stranded cassette was then purified by phenol extraction and ethanol precipitation. Complete digestion of the cassette with *Kpn*I and *Xho*I is carried out in a single incubation. The digested cassette is then purified by phenol and ether extractions and ultrafiltration in a Centricon 30 device (Amicon).

Ligation is carried out for 24 h at 16°C in 20 μl with approximately 0.1 μg of pU4b similarly digested with *Kpn*I and *Xho*I. The resulting pU4b derivative (an aliquot of the ligation) are directly electroporated into S17-1 *E.coli*. Aliquots of the transformation are plated on LB-tetracycline plates (after allowing 1 h for resistance expression) for complexity estimation and the remainder of the transformation is incubated overnight in 60 ml of LB-tetracycline. Plasmid pU4b derivatives were conjugated from *E.coli* S17-1 donors into *R.capsulatus* strain U71. The library is expressed by U71 transconjugants selected for by growth on RCV-tetracycline plates at 32°C.

Dope optimization

In computer simulations, various functions were used to optimize the 'nucleotide mixtures'. In this work, only five functional mutant sequences were obtained in the zero iteration. Given this small number of sequences and in order to conserve diversity, we elected to use the group probability (P_G) function because it retains all amino acids in the target set. When presented with a target set at one position, the program 'CyberDope' (provided courtesy of KAIROS Inc., Cambridge, MA, USA) goes through all integer nucleotide mixtures possible for a codon and evaluates for each mixture the value of P_G:

$$P_G = \prod_i P_D[i] \quad (1)$$

where $P_D[i]$ is the frequency of occurrence of the ith amino acid

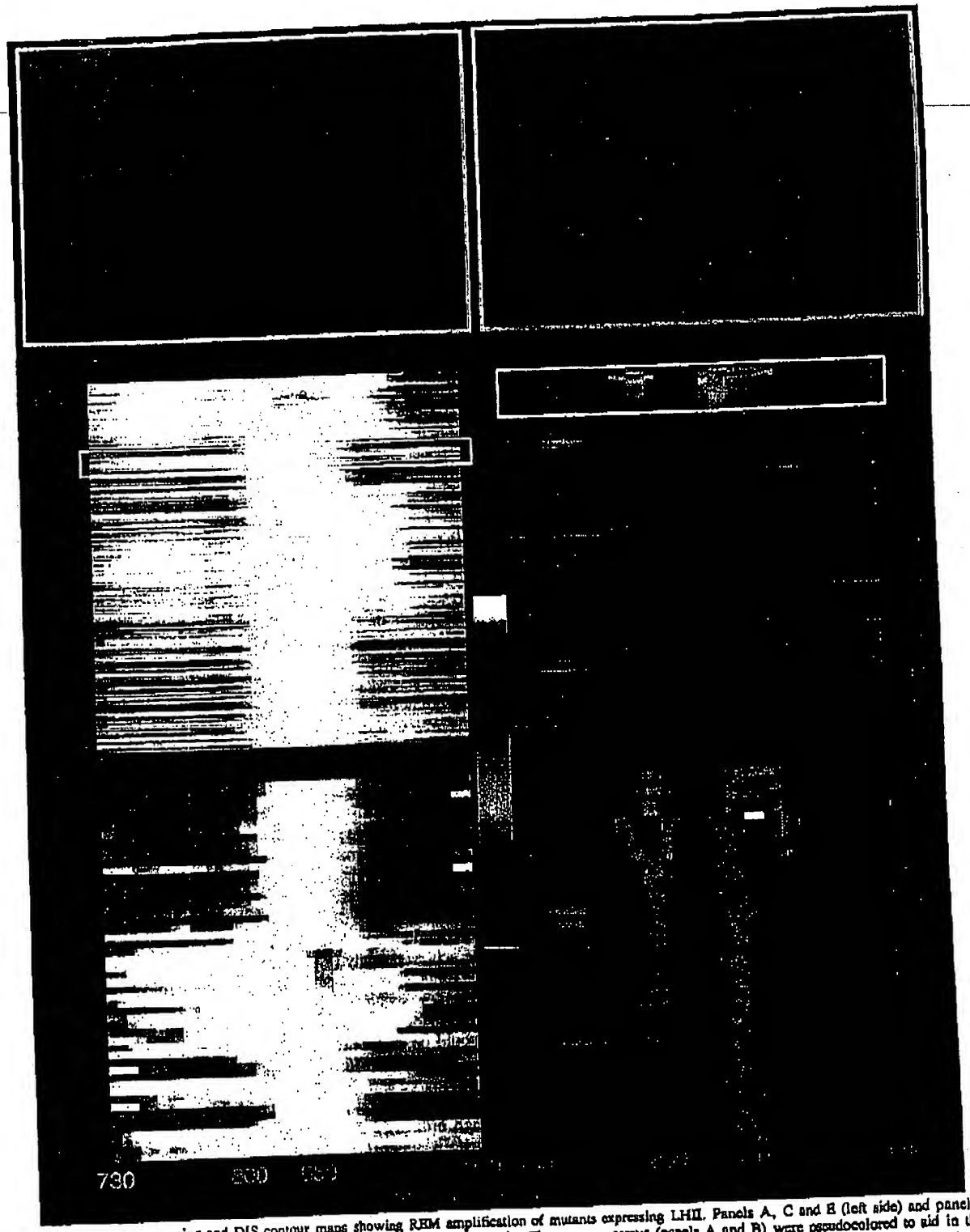

Fig. 2. Fluorescence screening and DIS contour maps showing REM amplification of mutants expressing LHII. Panels A, C and E (left side) and panels B, D and F (right side) correspond to the zero and first iterations of REM respectively. Fluorescence assays (panels A and B) were pseudocolored to aid in the visualization of relative levels of fluorescence according to the color bar shown below. These show brighter and more numerous fluorescent colonies in the first iteration of REM (panel B) than in the zero iteration of REM (panel A). DIS color contour maps (panels C and D) of the ground state absorption spectra of the same Petri dishes shown in panels A and B confirm the presence of mutants expressing Bchl-binding proteins. Each horizontal line represents the color coded absorption spectrum of a single colony; the color bar spans OD values of 0.0876–0.1177 for panels C and E and 0.0927–0.2088 for panels D and F. The spectra were sorted (Arkin and Youvan, in press) according to their absorption at 850 nm. Eight hundred and ninety-seven spectra from the zero iteration plate and 451 spectra from the first iteration plate are displayed in panels C and D respectively. Panels E and F highlight the spectra of 50 colonies from panels C and D (within the white boxes) respectively. Panel F shows many more colonies with spectra characteristic of LHII than found in panel E.

(in a target set of i amino acids) as encoded by a specific triplet dope. For the hypothetical target set mentioned above (Ala, Ser and Thr), any mixture not encoding a member of the target set (e.g. $P_D[Ala] = 0$) will cause P_G to be zero. The mixture with the highest value of P_G will be selected for the dope at that position. The doped sequence within the cassette used in the first iteration of REM was
[(G,T)(C,T)(C,G)][(A,G,T)(C,T)(C,G)][(C,T)(C,G)G][(C,T)(C,G)G][(A,G,T)(G,T)G][(C,T,G)(C,T,G)C]

Imaging spectroscopy

Colonies were imaged as spreads on RCV-tetracycline plates from the bacteria resuspended after conjugation. The most recent configuration of the digital imaging spectrophotometer has been described (Arkin and Youvan, in press). For the fluorescence images, the Petri dishes were illuminated with broad-band blue–green light and an 830 nm long pass filter was placed in front of the CCD lens to obtain radiometrically calibrated monochrome images which were linearly mapped to pseudocolors after establishing the low and high gray scale values for both images.

Results

The experimental complexity (i.e. number of independently generated clones) of the 'zero iteration' $[NN(G,C)]_6$ library was approximately 45 000. The theoretical complexity of such a library at the nucleotide level is calculated as 32^6 (1.1×10^9) because there are 32 possible [NN(G,C)] codons; the experimental complexity is only a small fraction of this number. Preliminary screening used fluorescence, (Yang and Youvan, 1988) which is indicative of LHII assembly, to rapidly identify mutants expressing LHII. Mutants are then more closely evaluated by ground state absorption measurements using DIS. We observed a low frequency of highly fluorescent colonies in the zero iteration of REM (ca. one positive mutant in 10 000 colonies screened). Relative to wild type absorption, DIS showed a decrease in the optical density at 800 and 858 nm for these few positives.

Because of their rarity, only five positives were obtained from the zero iteration of REM. Four of these five mutants fit the selection criterion of displaying significant absorbance at 858 nm and another, REM0.10, had an interesting phenotype. The five positives were repurified and sequenced (Table I). The composition of a first iteration cassette was calculated by the computer program 'CyberDope', which generates DNA dopes that maximize the overall probability of the target set. To add diversity to the target set, the wild type sequence was also included. Therefore, while not taking frequency of occurrence into account because of the small sample size, for the first doped position the target set is F, S, A, L. The output of CyberDope at the nucleotide level gave the codon [(G,T)(C,T)(C,G)], which encodes amino acids A, S, V (0.25 probability of occurrence for each) and F, L (0.12 probability of occurrence for each). Valine is unavoidably encoded by this dope because of the structure of the genetic code.

Figure 2 demonstrates the amplification properties of the REM methodology as assayed by digital imaging spectroscopy using both fluorescence emission and ground state absorption imagery. The first iteration of REM yields a 30-fold increase in the frequency of enhanced fluorescence mutants (Figure 2 A and B). As compared to zero iteration REM data, DIS analysis of the first iteration library shows both an increase in the percentage of positive mutants (i.e. throughput) and an increase in protein levels as determined by the intensity of the BChl absorption bands (Figure 2 and Table I).

Twelve positive mutants were sequenced from the first iteration REM library. All of these mutants express unique peptide sequences that differ from wild type. Two mutants (REM1.8 and REM1.9) show a 10 nm blue shift in the 858 nm band. These blue-shifted mutants have an inversion of the Pro–Trp motif found in all 29 sequences in the known phylogeny (Zuber, 1988) of β-subunits. This phenotype was first observed in the zero iteration library (mutant REM0.10), but now finds itself amplified in the first iteration. Note also that REM0.10 contains the same Pro–Trp motif inversion.

Discussion

To show that computer simulations were accurate in their prediction of an increased throughput of positives, an LHII gene was iteratively mutagenized at its six carboxy terminal residues. From the zero iteration (CCM) data, target sets of amino acids were defined. A computer encoded algorithm generated a doped oligonucleotide which best represented the target set at each mutagenized position. Expression of this new library (the first iteration of REM) revealed a substantial amplification in the throughput of pseudo wild type mutants. From the zero iteration library where roughly 10 000 colonies were screened to identify one positive, we can now conveniently identify a new positive by screening only about 300 colonies. This corresponds to a 30-fold increase in overall throughput, suggesting that mutating 18 sites of similar stringency would yield a 30^3 or 27 000-fold increase in throughput over random mutagenesis using $[NN(G,C)]_{18}$.

The altered proteins obtained by combinatorial mutagenesis are not necessarily trivial variations of the wild type sequence. An inversion of a completely conserved motif was observed in some mutants. Therefore, the sequence data indicate that REM does not recapitulate the known phylogeny. Mechanistically, the simultaneous (experimental) randomization of six sites in a protein may have no analogy in nature.

In this work, experimental evidence is given that REM allows an efficient search of sequence space by producing mutant libraries with increased frequencies of selected 'positives'. Due to the high stringency of the region chosen for mutagenesis, only a small sequence database was available for the construction of the first iteration dope. In systems where large complexities can be achieved easily (e.g. phage display libraries), more sites can be mutated at once and more positives isolated, giving a more complex sequence database. As a consequence, other dope optimizing equations (Youvan *et al.*, 1992) could be used which would be better suited to yield large increases in throughput. Alternatively, different short stretches of amino acids could be randomized and the zero iteration data from these libraries pooled to produce a first iteration dope mutagenizing many more sites than ordinarily possible with CCM.

It is important to make the connection between our algorithmically-based doping schemes and protein engineering projects where CCM is currently being used. REM decreases the fraction of null mutants in the population, therefore more sites can be simultaneously mutated. Model experiments on LHII can be used to optimize REM methodology, including the nucleotide doping equations. While DIS is limited to screening about 10^6 colonies, phage display libraries (Smith, 1985; Hoogenboom *et al.*, 1991; Kang *et al.*, 1991) can be used to select mutants from libraries with complexities exceeding 10^9. Based on our preliminary experiments, we expect greater phenotypic diversity

after one iteration of REM. This means that stronger 'binders' can be isolated, which is the fundamental goal of the phage display methodology. The use of CCM to introduce additional diversity in antibody libraries has already proven a useful approach (Barbas *et al.*, 1992) and may well be enhanced by the use of our mutagenesis scheme. REM is the first optimization technique that can be used to address this problem and explore sequence space in a mathematically rigorous fashion.

Acknowledgements

This work was supported by NIH GM42645, DOE 9102-025, DOE DE-FG02-90ER20019, and by a grant from the Human Frontiers Science Program. S.D. is partially supported by the fonds FCAR. E.R.G. is the recipient of an NIH Biotechnology Training Grant Award.

References

Arkin,A.P. and Youvan,D.C. (1992a) *Proc. Natl Acad. Sci. USA*, 89, 7811–7815.

Arkin,A.P. and Youvan,D.C. (1992b) *Biotechnology*, 10, 297–301.

Arkin,A.P. and Youvan,D.C. (1993) In Deisenhofer,J. and Norris,J.R. (eds), *The Photosynthetic Reaction Center*. Academic Press, New York, in press.

Arkin,A.P., Goldman,E.R., Robles,S.J., Coleman,W.J., Goddard,C.A., Yang,M.M. and Youvan,D.C. (1990) *Biotechnology*, 8, 746–749.

Barbas,C.F., Bain,J.D., Hoekstra,D.M. and Lerner,R.A. (1992) *Proc. Natl Acad. Sci. USA*, 89, 4457–4461.

Beaudry,A.A. and Joyce,G.F. (1992) *Science*, 257, 635–641.

Goldman,E.R. and Youvan,D.C. (1992) *Biotechnology*, 10, 1557–1561.

Hoogenboom,H.R., Griffiths,A.D., Johnson,K.S., Chiswell,D.J., Hudson,P. and Winter,G. (1991) *Nucleic Acids Res.*, 19, 4133–4137.

Kang,A.S., Barbas,C.F., Janda,K.D., Benkovic,S.J. and Lerner,R.A. (1991) *Proc. Natl Acad. Sci. USA*, 88, 4363–4366.

Oliphant,A.R., Nussbaum,A.L. and Struhl,K. (1986) *Gene*, 44, 177–183.

Reidhaar-Olson,J.F., Bowie,J.U., Breyer,R.M., Hu,J.C., Knight,K.L., Lim,W.A., Mossing,M.C., Parnell,D.A., Shoemaker,K.R. and Sauer,R.T. (1991) *Methods Enzymol.*, 208, 564–587.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sjostrom,M. and Wold,A. (1985) *J. Mol. Evol.*, 22, 272–277.

Smith,G.P. (1985) *Science*, 228, 1315–1317.

Yang,M.M. and Youvan,D.C. (1988) *Biotechnology*, 6, 939–942.

Youvan,D.C. (1991) *Trends Biochem. Sci.*, 16, 145–149.

Youvan,D.C. and Ismail,S. (1985) *Proc. Natl Acad. Sci. USA*, 82, 63–67.

Youvan,D.C., Ismail,S. and Bylina,E.J. (1985) *Gene*, 38, 19–30.

Youvan,D.C., Arkin,A.P. and Yang,M.M. (1992) In Muenner,R. and Munderick,B. (eds), *Parallel Problem Solving from Nature*, 2. Elsevier Publishing Co., Amsterdam, pp. 401–410.

Zuber,H. (1990) In Drews,G. and Dawes,E.A. (eds), *Molecular Biology of Membrane-bound Complexes in Phototrophic Bacteria*. Plenum Press, New York, pp. 161–180.

Received on November 6, 1992; revised on December 17, 1992; accepted on December 22, 1992

Combinatorial Cassette Mutagenesis as a Probe of the Informational Content of Protein Sequences

JOHN F. REIDHAAR-OLSON AND ROBERT T. SAUER

A method of combinatorial cassette mutagenesis was designed to readily determine the informational content of individual residues in protein sequences. The technique consists of simultaneously randomizing two or three positions by oligonucleotide cassette mutagenesis, selecting for functional protein, and then sequencing to determine the spectrum of allowable substitutions at each position. Repeated application of this method to the dimer interface of the DNA-binding domain of λ repressor reveals that the number and type of substitutions allowed at each position are extremely variable. At some positions only one or two residues are functionally acceptable; at other positions a wide range of residues and residue types are tolerated. The number of substitutions allowed at each position roughly correlates with the solvent accessibility of the wild-type side chain.

IT HAS BEEN MORE THAN 20 YEARS SINCE ANFINSEN AND HIS colleagues showed that the sequence of a protein contains all of the information necessary to specify the three-dimensional structure (*1*). However, the general problem of predicting protein structure from sequence remains unsolved. Part of the difficulty may stem from the complexity of protein structures. Although some 200 protein structures are known, no rules have emerged that allow structure to be related to sequence in any simple fashion (*2*). The problem is further complicated by the nonuniformity of the structural information encoded in protein sequences. Some residue positions are important, and changes at these positions can tip the balance between folding and unfolding (*3–7*). Other residues are relatively unimportant in a structural sense and a wide range of substitutions or modifications can be tolerated at these positions (*3, 7–9*).

If only a fraction of the residues in a protein sequence contribute significantly to the stability of the folded structure, then it becomes important to be able to identify these residues. We now describe the results of genetic studies that allow the importance of individual residues in protein sequences to be rapidly determined. Specifically, we determine the spectrum of functionally acceptable substitutions at residue positions near the dimer interface of the NH_2-terminal domain of phage lambda (λ) repressor (*10*). The NH_2-terminal domain binds to operator DNA as a dimer, with dimerization mediated by hydrophobic packing of α helix 5 of one monomer against α helix 5′ of the other monomer (*11*) (Fig. 1, A and B). Without helix 5 there are no contacts between the subunits (Fig. 1C). By applying combinatorial cassette mutagenesis to the helix 5 region, we find that the number and spectrum of allowable substitutions within helix 5 are extremely variable from residue to residue. In most cases, this variability can be rationalized in terms of the fractional solvent accessibility of the wild-type side chain.

General strategy. For our studies, we used a plasmid-borne gene that encodes a functional, operator-binding fragment (residues 1–102) of λ repressor (*12*). The binding of the 1–102 fragment to operator DNA depends on dimerization which, in turn, depends on the helix 5–helix 5′ packing interactions (*11, 13*). Thus, if a 1–102 protein retains normal operator-binding properties, we can infer that it is able to dimerize normally.

Mutagenesis of the helix 5 region was performed by a combinatorial cassette procedure. One example of this method, in which codons 85 and 88 are mutagenized, is illustrated in Fig. 2. On the top strand, the mutagenized codons are synthesized with equal mixtures of all four bases in the first two codon positions and an equal mixture of G and C in the third position. The resulting population of base combinations will include codons for each of the 20 naturally occurring amino acids at each of the mutagenized residue positions. On the bottom strand, inosine is inserted at each randomized position because it is able to pair with each of the four conventional bases (*14*). The two strands are then annealed and the mutagenic cassette is ligated into a purified plasmid backbone.

To identify plasmids encoding functional protein, we selected transformants for plasmid-encoded resistance to ampicillin and for resistance to killing by *cI*⁻ derivatives of phage λ. The latter selection requires that the cell express 1–102 protein that is active in operator binding (*15*). For each mutagenesis experiment, many independent transformants were chosen, single-stranded plasmid DNA was purified, and the relevant region of the 1–102 gene was sequenced. The resulting set of sequences provides a list of functionally acceptable helix 5 residues.

Substitutions in the helix 5 region. In separate experiments with different mutagenic cassettes, the codons for helix 5 residues 85 and 88; 86 and 89; 90 and 91; 84, 87, and 88; and 84, 87, and 91 were mutagenized, and genes encoding active 1–102 proteins were selected. In some cases, the survival frequency was low. For example, only 17 of 60,000 transformants passed the selection after randomization of codons 84, 87, and 88. In this case, each active candidate was sequenced. By contrast, 1,200 of 50,000 transformants passed the selection in the mutagenesis of positions 86 and 89 (*16*). In this case, we picked 50 candidates for sequence analysis. Overall, 150 active genes were sequenced (Table 1). In addition, we sequenced

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

approximately 40 genes that had been mutagenized, but not subjected to a functional selection. These serve as controls for the efficiency of mutagenesis and also provide examples of helix 5 mutations that result in inactive 1–102 proteins (Table 1).

Many of the active sequences contain at least two residue changes compared to wild type. In principle, some of these changes could be compensatory; for example, residue X might be functionally allowed at position 85 only in combination with residue Z at position 88. This cannot be generally true, however, because most residue changes at one position were recovered in combination with several different changes at the other position or positions. It is therefore likely that most substitutions that are functionally acceptable in multiply mutant backgrounds would also be allowed as single substitutions. In Fig. 3, we show the spectrum of functionally acceptable substitutions at residue positions 84 to 91.

From the list of allowed substitutions, several conclusions may be drawn concerning the structural requirements at various positions in helix 5. We now consider these residue positions in order of decreasing "informational content," where this term is roughly defined as a value that decreases as the number of allowed substitutions increases. Thus, the informational content of a residue position is highest if only the wild-type amino acid is allowed and is lowest if each of the 20 naturally occurring amino acids is allowed.

Positions 84 and 87 in particular stand out as having a high informational content. Ile appears to be the only acceptable residue at position 84. Both Met and Leu are residues of similar size and hydrophobicity, and are the only two residues that appear to be functional at position 87. The side chains of Ile^{84} and Met^{87} form a major part of the helix-helix packing interaction at the dimer interface, where Ile^{84} of one subunit packs against $Met^{87'}$ of the other subunit, and vice versa (Fig. 4). This cluster of four residues also contacts the globular portions of the domain. Solvent accessibility calculations by the method of Lee and Richards (*17*) show that the Ile^{84} and Met^{87} side chains are almost completely buried (92 to 98 percent solvent inaccessible) in the structure of the dimer. We assume that replacement of Ile^{84} or Met^{87} with smaller side chains would diminish dimerization because hydrophobic and van der Waals interactions would be lost. In fact, mutant repressors containing Ser^{84} or Thr^{87} are defective in dimerization (*13, 18*). Replacing Ile^{84} or Met^{87} with larger residues would also be expected to be detrimental because substantial structural rearrangements would be required to accommodate larger side chains.

Seven residues (Leu, Ile, Val, Thr, Cys, Ser, and Ala) are functionally acceptable at position 91. Aromatic residues, charged residues, and strongly hydrophilic residues are not found. The wild-type Val side chain is partially buried in the dimer structure, with the Cγ2 methyl group packing against the Cδ1 methyl group of the $Ile^{84'}$ side chain. Although some of the acceptable substitutions such as Ile and Thr could make equivalent packing contacts, others such as Ala and Ser could not.

Nine residues (Trp, His, Met, Gln, Leu, Val, Ser, Gly, and Ala) are acceptable at position 90. There is a surprisingly large range in both the acceptable size and hydrophilicity of these side chains. This is especially true as the Cβ methyl group of the wild-type Ala is almost completely buried in the structure of the dimer and, at first glance, it would appear that larger side chains could not be accommodated. However, the inaccessibility of the Cβ methyl group of Ala^{90} is largely caused by the $Lys^{67'}$ side chain, which packs against it. By rotating the $Lys^{67'}$ side chain away, we were able to introduce a Trp^{90} side chain by model-building without steric clashes. Rotation of the $Lys^{67'}$ side chain away from Ala^{90} should not be energetically costly and, in fact, is observed in crystals of the NH_2-terminal domain bound to operator DNA (*19*).

Nine different residues (Trp, Tyr, Phe, Met, Ile, Val, Cys, Ser, and Ala) are functionally acceptable at position 88. There are large variations in the sizes and volumes of the acceptable side chains, although most are relatively hydrophobic. Charged residues and other strongly hydrophilic residues are not observed. In the wild-type dimer (*11*), the aromatic ring of Tyr^{88} stacks against the ring of $Tyr^{88'}$. The side chains of Trp, Phe, Met, Ile, and Val could probably form some type of packing interaction at this position, although those of Ala and Ser could not. It is known that the presence of Cys at position 88 allows a stable Cys^{88}–$Cys^{88'}$ disulfide bond, which links the monomers in a conformation that is active in operator binding (*20*).

Positions 85, 86, and 89 show considerable variability. At each of these positions, 13 different amino acids were found to function. At positions 85 and 86, aromatic, hydrophobic, polar, and charged residues are all acceptable. At position 89, aromatic residues were not represented, but each of the remaining classes was observed. In

Table 1. Sequences for the helix 5 region of active and inactive mutants obtained by combinatorial cassette mutagenesis. Active mutants are resistant to phage λKH54; these are grouped by cassette, with the wild-type sequence at the top of each group and randomized positions in boldface. Asterisks indicate sequences of mutants obtained in the absence of a functional selection. The activity of these mutants was subsequently determined by a screen. Numbers next to sequences indicate the number of times particular mutant sequences were obtained. Numbers at the tops of the columns indicate amino acid positions. The one-letter abbreviations for the amino acids are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Active

85 90	85 90	85 90	85 90
IYEMYEAV	-L--V---	--Y--S--	--E--L--*
I--MF--- 2	-L--M---	--Y--T--	--E--E--
I--MY--- 4	-I--F---	--W--A--	--E--G--*
I-AMA---	-I--A---	--W--C--	--G--E--
I-DMY---	-I--S---	--W--Q--	
I--MA--- 3	-Q--Y---	--W--S-- 3	IYEMYEAV
I--MI---	-S--Y---	--W--K--*	------WA 4
I--LF---	-S--W---	--W--D--	------WL*
I--LW---	-T--Y---	--W--D--*	------WS
	-T--W---	--W--E--	------HV
IYEMYEAV	-T--A---	--W--G-- 3	------AV 6
I--M---V	-R--F---	--A--E-- 3	------AC 2
I--M---T	-R--W--- 4	-SM--A--	------AL 4
I--L---T	-R--A--- 3	--M--S-- 3	------AI
	-R--C---	--L--S-- 2	------AT 2
IYEMYEAV	-E--F---	--L--T--	------VA
-Y--F---	-G--Y---	--L--E--	------VC 2
-W--W--- 2		--L--E--*	------MA
-W--A---	IYEMYEAV	--I--S--	------LA*
-A--Y---	--F--M--	--Q--S--	------QV
-V--Y--- 2	--F--L--	--S--Q--	------QT
-V--A--- 3	--F--I--	--S--E--	------SV 3
-C--F--- 2	--F--Q--	--T--E--	------SC 2
-C--A---	--F--S-- 3	--D--L--	------SL 5
-L--F---	--F--T--	--D--Q--	------ST
-L--W---	--F--R--	--D--D--*	------GC
-L--A---	--F--E--	--E--L-- 2	------GI
			------GT

Inactive

85 90	85 90	85 90	85 90
A--VA---*	P--DS---*	Q--CS---*	------RN*
P--PL---*	R--TR---*	A--TP---*	------EA*
P--TN---*	T--TV---*	S--TK---*	------KV*
R--NP---*	R--VI---*	T--LN---*	------VM*
P--LL---*	L--PL---*	A--SL---*	------PA*
A--IL---*	I--LL---*	R--WS---*	------NQ*
T--KP---*	K-AIV---*	------PR*	------ME*
Q--RV---*	C--YT---*	------PP*5	------AY*
K-DVR---*			

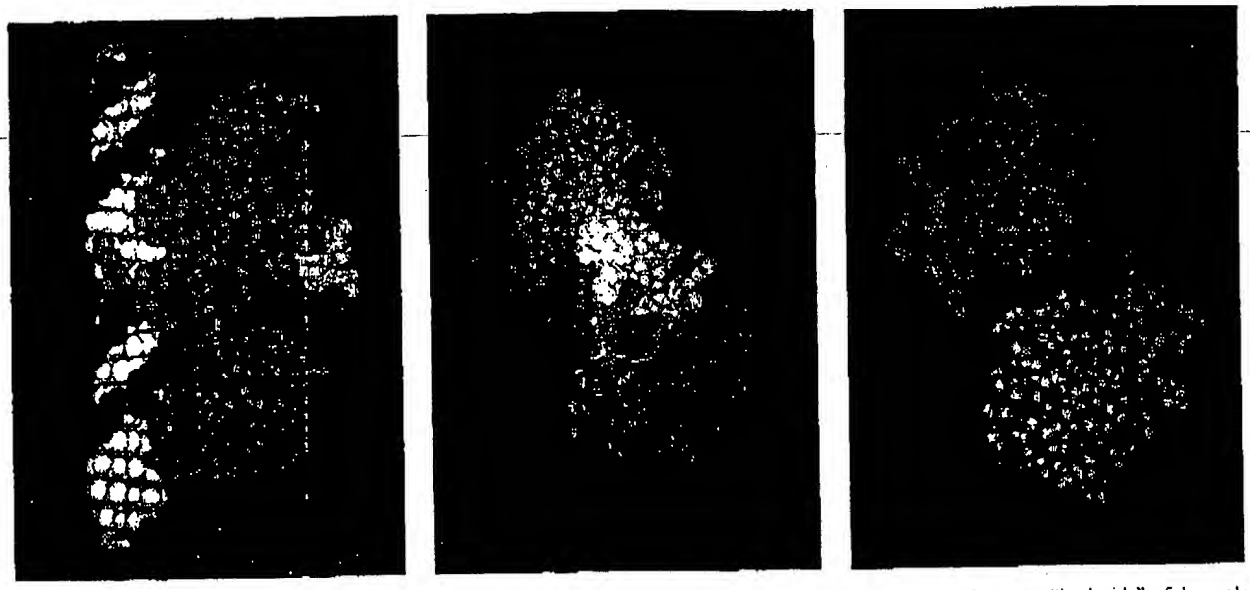

Fig. 1. Three views of the DNA-binding domain of λ repressor, showing the role of helix 5 in dimerization. (**A**) Proposed complex of repressor dimer with operator DNA (*11*). Helix 5 of each monomer is colored more lightly than the globular portion of that monomer. (**B**) Free repressor dimer, rotated 90° from the view in (A), to show the "back side" of the molecule. (**C**) Dimer with helix 5 of each monomer removed. This view illustrates the role helix 5 plays in mediating dimerization (*26*).

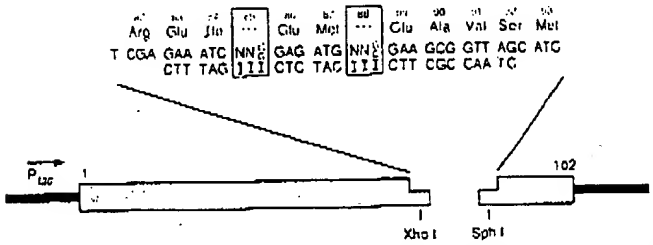

Fig. 2. Schematic diagram showing the combinatorial cassette mutagenesis procedure. At positions indicated as N, an equal mixture of A, G, C, and T was used, during oligonucleotide synthesis. At positions indicated as I, inosine was used. After synthesis, the oligonucleotides were phosphorylated, annealed, and ligated into the Xho I–Sph I backbone of plasmid pJO103. Plasmid pJO103 is an M13 origin plasmid with the 1–102 gene under control of a *tac* promoter; the region of the 1–102 gene encoding residues 82–93 (the small Xho I–Sph I fragment) is replaced by an unrelated 1.9-kb Xho I–Sph I "stuffer" fragment. Ligated DNA was transformed into *Escherichia coli* strain X90 F′*lacI*^{Q} cells (*27*), and ampicillin-resistant colonies were selected in the presence or absence of phage λKH54. Candidates that survived the selection were cross-streaked against a series of virulent derivatives of phage λ to confirm their immunity properties [strains and methods are described in (*21*)]. Single-stranded plasmid DNA was purified from an M13RV1 transducing lysate as described (*28*), and DNA sequences were determined by the dideoxy method (*29*).

the wild-type dimer, the side chains of Tyr^{85}, Glu^{86}, and Glu^{89} are relatively solvent accessible.

Several amino acids are significantly underrepresented among the active sequences. For example, Pro is never found. This cannot be an artifact of our mutagenesis procedure because Pro is frequently observed among the unselected mutant sequences (Table 1). We conclude that Pro is not found among the functional sequences because it is selected against; its presence would presumably disrupt the α-helical structure and thereby the helix-helix packing at the dimer interface.

His, Asn, and Lys are also underrepresented among the functional helix 5 sequences. These residues are presumably not acceptable at positions 84 and 87, where the informational content is extremely high, and may not be acceptable at positions 88 and 91, where the functional substitutions are generally hydrophobic in character. The acceptability of these residues at positions such as 85 and 86 is difficult to assess from our experiments because the codons for these residues are present at reasonably low frequencies even among the unselected sequences. In these cases, we probably have not sequenced a large enough number of candidates to be confident that all acceptable substitutions have been identified. In fact, data from reversion studies (*21*) and suppressed amber studies (*22*) show that His^{85} and Lys^{86} are acceptable substitutions in the context of the intact λ repressor molecule.

Informational content and protein structure. We have combined an efficient combinatorial mutagenesis procedure and a functional selection to probe the informational content of the eight residues that form the major part of the dimerization interface of the NH_2-terminal, operator-binding domain of λ repressor. At two of these eight residue positions, the functionally acceptable choices are highly restricted. For example, we analyzed 17 functional genes in which codon 84 had been randomized and recovered the wild-type residue, Ile, in every case. This is clearly a position of high

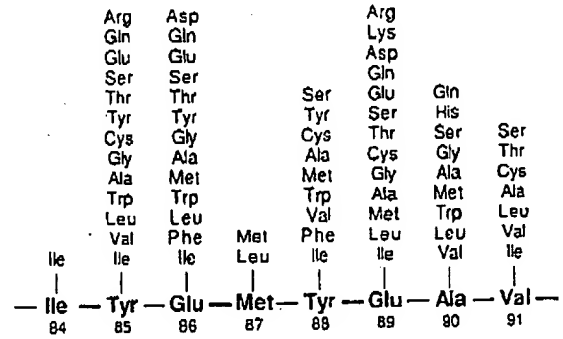

Fig. 3. Functionally acceptable residues in the helix 5 region. The amino acids are listed from top to bottom in order of increasing hydrophobicity according to the scale of Eisenberg *et al.* (*30*).

Fig. 4. Helix 5 residues high in informational content. The two isolated helix 5 regions of the protein are shown in green and blue. Ile^{84} and Met^{87} from the green helix are shown in yellow; Ile^{84} and Met^{87} from the blue helix are shown in red.

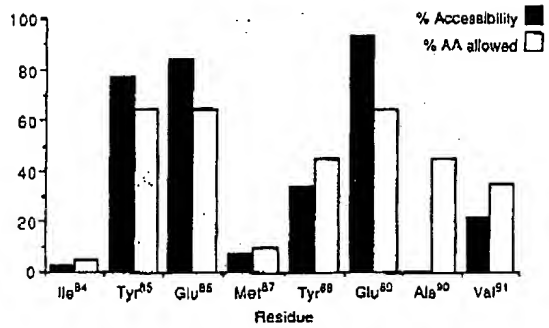

Fig. 5. Correlation between the solvent accessibility and the number of functionally acceptable substitutions. Hatched bars indicate the percentage of the 20 naturally occurring amino acids that are functionally acceptable at a residue position. Black bars indicate the fractional solvent accessibility of the wild-type side chain in the dimer. Solvent accessibilities for the NH_2-terminal domain dimer (*11*) were computed using a 1.4 Å probe by the method of Lee and Richards (*17*). Fractional accessibilities were obtained by dividing by the appropriate side chain accessibilities calculated for the monomer. The fractional accessibilities change only slightly if the side chain accessibilities in the reference tripeptide Ala-X-Ala (*17*) are used instead as the reference state.

informational content. The informational content is also high at position 87, where Met and Leu are the only acceptable residues. By contrast, the remaining positions have moderate to low informational contents. For example, among 38 functional genes in which codon 85 had been randomized, the wild-type residue was recovered only once, and 12 other residues, differing in size and chemical properties, were recovered in the remaining cases. This is clearly a position of low informational content. It is striking that most of the structural determinants of dimerization in this eight-residue segment reside in two residues only. The remaining positions are surprisingly tolerant of a wide range of substitutions. If this high level of tolerance is generally true of protein sequences, then the problem of understanding and predicting structure may rest largely on the ability to identify those few residues that are crucial.

The positional variability of the informational content in helix 5 can, in general, be rationlized in terms of the solvent accessibility of the wild-type residues in the crystal structure (*11*). There is a rough correlation between the number of acceptable substitutions and the fractional extent to which the wild-type side chain is solvent accessible (Fig. 5). At exposed surface positions such as 85, 86, and 89, we find that many different residues and residue types can be functionally accommodated. By contrast, at positions such as 84 and 87, where the wild-type side chain is almost completely buried, we find that the functionally acceptable residue choices are extremely restricted. There is one apparent exception to the simple rule that buried residues are high in informational content. Ala^{90} is inaccessible to solvent in the crystal structure, and yet we find that many substitutions are allowed at this position. However, the inaccessibility of the Ala^{90} side chain to solvent is not due to close packing at the dimer interface, but rather to an interaction with a nearby surface side chain. This side chain can presumably move to allow larger side chains to be accommodated at position 90. Examples of this type demonstrate the need to distinguish between two types of buried side chains: those that can become exposed by relatively minor rearrangement of other side chains, and those that are tightly packed in the hydrophobic core.

There is no reason to assume that there should always be a strict correlation between the solvent accessibility of a residue and the structural informational content of that position. For one thing, the chemical properties of the 20 amino acids are not related in any simple linear fashion. Moreover, the structural importance of some residues in proteins almost certainly stems from interactions other than simple hydrophobic packing. Nevertheless, the closely packed nature of protein interiors (*23*) provides a simple molecular explanation for the structural importance of buried residues, and destabilizing mutations are commonly found to affect hydrophobic core residues (*3*–*7*). By contrast, missense mutations or chemical modifications that affect surface residues are often found to have little or no influence on protein stability (*3*, *7*, *8*). Thus, it is reasonable that solvent accessibility should be an extremely important determinant of the informational content of a residue position.

Our overall strategy for rapidly probing informational content should be broadly applicable to a wide range of protein structure-function problems in systems where genetic selections or screens can be devised. The method consists of three basic elements: (i) the use of cassette mutagenesis to introduce extremely high levels of targeted random mutagenesis; (ii) the use of a functional selection to identify genes encoding active proteins; and (iii) the use of rapid DNA sequencing methods to determine the spectrum of functionally acceptable residues in a relatively large number of candidates. Our method of combinatorial cassette mutagenesis (Fig. 2) allows several residue positions to be mutagenized at the same time and, in principle, generates a mutant population in which each of the 20 amino acids is represented at each mutagenized position (*24*). When two or three codons are mutagenized at the same time, the entire analysis is able to proceed more rapidly. Moreover, at this level of mutagenesis most two-residue and three-residue combinations should be present in the mutagenized population and should be recovered if they result in a functional protein. In our study of the packing of the 84 and 87 side chains, we recovered only two (Ile^{84} with Met^{87} and Ile^{84} with Leu^{87}) of the 400 possible residue combinations. Thus, because both positions were mutagenized in the same experiment, we are able to conclude that there are not significantly different ways of packing the dimer interface.

In principle, data like that shown in Fig. 3 could be generated for an entire protein sequence, and additional experiments could be devised to determine whether the positions of high informational content were important for structure or function. For proteins of unknown structure, such data might be quite useful for structural predictions. First, current predictive algorithms could be applied to the family of related sequences generated by our method, as each of these sequences is able to form the same basic structure. Second, because of their fundamental repeats, α-helical and β-strand regions might be recognized by characteristic patterns of high and low informational content. Third, the positions of highest structural informational content should include the residues involved in

formation of the hydrophobic core of the protein. This information might prove useful in combination with the tertiary template ideas recently proposed (*25*).

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. T. E. Creighton, *Proteins: Structures and Molecular Properties* (Freeman, New York, 1983), chap. 6.
3. M. H. Hecht, H. C. M. Nelson, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2676 (1983); M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *ibid.* **81**, 5685 (1984); M. H. Hecht, K. M. Hehir, H. C. M. Nelson, J. M. Sturtevant, R. T. Sauer, *J. Cell. Biochem.* **29**, 217 (1985).
4. D. Shortle and B. Lin, *Genetics* **110**, 539 (1985); D. Shortle and A. K. Meeker, *Proteins* **1**, 81 (1986).
5. A. Pakula, V. Young, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8829 (1986).
6. T. Alber, S. Dao-pin, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
7. G. Fermi and M. F. Perutz, *Haemoglobin and Myoglobin* (Clarendon, Oxford, 1981).
8. M. Hollacker and T. E. Creighton, *Biochem. Biophys. Acta* **701**, 395 (1982).
9. J. H. Miller, in *The Operon*, J. H. Miller and W. S. Reznikoff, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1978), pp. 31–88.
10. λ repressor consists of an NH_2-terminal and COOH-terminal domain; each domain contains an independent set of dimerization contacts [C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1608 (1979); C. O. Pabo, thesis, Harvard University (1980)].
11. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982).
12. The NH_2-terminal domain of λ repressor is minimally comprised of residues 1–92. This is the fragment for which the crystal structure is known. In our studies we use a slightly larger NH_2-terminal fragment consisting of repressor residues 1–102. Nuclear magnetic resonance studies show that the 1–92 and 1–102 fragments have the same basic structure and similar dimerization properties (*13*). The operator binding properties of the 1–92 and 1–102 fragments are also extremely similar (*20*).
13. M. A. Weiss, C. O. Pabo, M. Karplus, R. T. Sauer, *Biochemistry* **26**, 897 (1987); M. A. Weiss, M. Karplus, R. T. Sauer, *ibid.*, p. 890.
14. F. H. Martin, M. M. Castro, F. Aboul-ela, I. Tinoco, *Nucleic Acids Res.* **13**, 8927 (1985).
15. In control experiments, we found that cells containing at least 5 to 10 percent of the wild-type activity could survive the phage selection. This figure was determined by Western analysis of lysates of cells containing the wild-type 1–102 gene under control of the inducible *lac* promoter. When expression of 1–102 was induced to a level sufficient to make cells resistant to phage λKH54, the intracellular level of 1–102 was 5 to 10 percent of that produced from the *tac* promoter.
16. The survival frequency can be somewhat misleading, as some cells containing functional genes do not survive the selection. For example, by screening the unselected candidates in the mutagenesis of positions 86 and 89, we found that approximately 20 percent of the cells contained active protein. By contrast, only 2.4 percent of the cells survived the selection in this experiment.
17. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971); in our studies, accessible surface areas were calculated with the use of the ACCESS program written by T. Richmond.
18. F. S. Gimble and R. T. Sauer, *J. Mol. Biol.*, in press.
19. S. Jordan and C. Pabo, personal communication.
20. R. T. Sauer *et al.*, *Biochemistry* **25**, 5992 (1986); C. O. Pabo and E. G. Sucharek, *ibid.* **25**, 5987 (1986).
21. M. H. Hecht and R. T. Sauer, *J. Mol. Biol.* **186**, 53 (1985).
22. H. C. M. Nelson, M. H. Hecht, R. T. Sauer, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 441 (1983).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974); C. Chothia, *Nature* **254**, 304 (1975).
24. There are several methods, in addition to our own, that could be used for combinatorial cassette mutagenesis: A. R. Oliphant, A. L. Nussbaum, K. Struhl, *Gene* **44**, 177 (1986); K. M. Derbyshire, J. J. Salvo, N. D. F. Grindley, *ibid.* **46**, 145 (1986). The use of inosines, in our method, is convenient and efficient, but does lead to some bias in the frequency at which each base is recovered. For example, among the unselected candidates in the mutagenesis of positions 84, 87, and 88, the frequencies at which the four bases were recovered at the mutagenized positions were: 23 percent A, 35 percent C, 20 percent G, and 23 percent T. The methods cited above avoid a pairing bias by performing enzymatic second-strand synthesis.
25. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
26. These graphics were produced with the Promodeler molecular graphics program (New England BioGraphics). Coordinates were provided by C. O. Pabo.
27. D. Hanahan, *J. Mol. Biol.* **166**, 557 (1983).
28. A. K. Vershon, K. Blackmer, R. T. Sauer, in *Protein Engineering: Applications in Science, Medicine, and Industry*, M. Inouye and R. Sarma, Eds. (Academic Press, Orlando, FL, 1986), pp. 243–256.
29. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
30. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, W. Wilcox, *Faraday Symp. Chem. Soc.* **17**, 109 (1982).

31. We thank R. Breyer for providing plasmids and antibodies required for the work, D. Vershon for suggesting the use of inosines for the combinatorial mutagenesis method, and J. Bowie for pointing out the potential uses of the method in structural prediction. We also thank C. Pabo and S. Jordan for providing the coordinates of the NH_2-terminal domain and its complex with operator DNA, and thank G. Quigley for help with the accessibility calculations. This work was supported by NIH grant AI-15706 and by a predoctoral grant (to J.R.-O.) from the National Science Foundation.

3 March 1988; accepted 18 May 1988